

ACCURACY OF EMPIRICAL PROJECTIONS OF HIGH-DIMENSIONAL GAUSSIAN MATRICES

BY ANGELIKA ROHDE

Universität Hamburg

Let $X = C + E$ with a deterministic matrix $C \in \mathbb{R}^{M \times M}$ and E some centered Gaussian $M \times M$ -matrix whose entries are independent with variance σ^2 . In the present work, the accuracy of reduced-rank projections of X is studied. Non-asymptotic universal upper and lower bounds are derived, and favorable and unfavorable prototypes of matrices C in terms of the accuracy of approximation are characterized. The approach does not involve analytic perturbation theory of linear operators and allows for multiplicities in the singular value spectrum. Our main result is some general non-asymptotic upper bound on the accuracy of approximation which involves explicitly the singular values of C , and which is shown to be sharp in various regimes of C . The results are accompanied by lower bounds under diverse assumptions. Consequences on statistical estimation problems, in particular in the recent area of low-rank matrix recovery, are discussed.

1. Introduction. As a consequence of the [Bai and Yin \(1993\)](#) law, the maximal singular value $\lambda_{\max}(E)$ of an iid standard Gaussian $M \times M$ -matrix E is equal to $2\sqrt{M}(1 + o(1))$ a.s. Since in addition the sequence $\lambda_{\max}(E)/\sqrt{M}$ is uniformly integrable ([Johnson and Lindenstrauss \(2001\)](#), Chapter 8, Theorem 2.4), the corresponding bound holds in expectation as well. Similarly, $\mathbf{E}\lambda_{\max}(E)^2 = 4M(1 + o(1))$. Let $\|\cdot\|_{S_2}$ denote the Hilbert-Schmidt or Frobenius norm. Define $\hat{\pi}_1$ to be the orthogonal projection matrix onto the one-dimensional subspace of \mathbb{R}^M maximizing $\|\hat{\pi}_1 E\|_{S_2}^2$ over all one-dimensional orthogonal projections $\tilde{\pi}_1$. Rewriting $\lambda_{\max}(E)^2 = \|\hat{\pi}_1 E\|_{S_2}^2$ yields

$$(1.1) \quad \mathbf{E}\|\hat{\pi}_1 E\|_{S_2}^2 = 4M(1 + o(1)).$$

In contrast, $\mathbf{E}\|\tilde{\pi}_1 E\|_{S_2}^2 = M$ for every fixed $\tilde{\pi}_1$. Thus, replacing one single projection by the supremum over all projections increases the Hilbert-Schmidt norm by a positive factor:

$$(1.2) \quad \mathbf{E}\|\hat{\pi}_1 E\|_{S_2}^2 - \mathbf{E}\|\tilde{\pi}_1 E\|_{S_2}^2 = 3(1 + o(1))M.$$

This effect raises the question about the accuracy for empirical reduced-rank projections in general. Consider the model

$$(1.3) \quad X = C + E$$

with a deterministic matrix $C \in \mathbb{R}^{M \times M}$ and E some centered Gaussian $M \times M$ -matrix whose entries are independent with variance σ^2 . Here and subsequently,

let

$$(1.4) \quad \hat{\pi}_r := \underset{\tilde{\pi}_r \in \mathcal{S}_{M,r}}{\operatorname{Argmax}} \|\tilde{\pi}_r X\|_{S_2}^2 \quad \text{and} \quad \pi_r \in \underset{\tilde{\pi}_r \in \mathcal{S}_{M,r}}{\operatorname{Argmax}} \mathbf{E} \|\tilde{\pi}_r X\|_{S_2}^2$$

with $\mathcal{S}_{M,r}$ denoting the set of all $M \times M$ -matrices representing orthogonal projections onto r -dimensional linear subspaces of \mathbb{R}^M . How close is $\mathbf{E} \|\hat{\pi}_r X\|_{S_2}^2$ to its deterministic counterpart $\mathbf{E} \|\pi_r X\|_{S_2}^2 = \|\pi_r C\|_{S_2}^2 + \sigma^2 r M$ if the Gaussian matrix X is not centered? For every fixed $M \in \mathbb{N}$ and $\sigma^2 > 0$, the following questions are natural:

- (A) Does there exist some favorable matrix $C = \mathbf{E}X$ for which the accuracy of approximation $\mathbf{E} \|\hat{\pi}_r X\|_{S_2}^2 - \mathbf{E} \|\pi_r X\|_{S_2}^2$ improves over the situation described in (1.2)?
- (B) Does there exist for any arbitrarily large real number c some unfavorable matrix $C(c)$ such that $\mathbf{E} \|\hat{\pi}_r X\|_{S_2}^2 - \mathbf{E} \|\pi_r X\|_{S_2}^2 \geq c$?

Based on the random variable $X = C + E$ within model (1.3), denote the difference by

$$(1.5) \quad \delta_{C,M,\sigma^2,r} := \mathbf{E} \|\hat{\pi}_r X\|_{S_2}^2 - \mathbf{E} \|\pi_r X\|_{S_2}^2,$$

which, in terms of singular values, is equal to

$$\sum_{i=1}^r \mathbf{E} \left(\hat{\lambda}_i^2 - \lambda_i^2 - \sigma^2 M \right),$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_M$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ denote the singular values of X and C , respectively. The goal in the present article is to study this quantity $\delta_{C,M,\sigma^2,r}$, to derive universal upper and lower bounds, and to characterize favorable and unfavorable types of matrices C in terms of the accuracy of approximation $\delta_{C,M,\sigma^2,r}$.

The motivation for considering this problem is two-fold. First of all, as

$$\mathbf{E} \|\hat{\pi}_r X\|_{S_2}^2 = \left(\mathbf{E} \|\hat{\pi}_r X\|_{S_2}^2 - \mathbf{E} \|\pi_r X\|_{S_2}^2 \right) + \|\pi_r C\|_{S_2}^2 + \sigma^2 r M$$

and $\mathbf{E} \|\hat{\pi}_r X\|_{S_2}^2 - \mathbf{E} \|\pi_r X\|_{S_2}^2 = \delta_{C,M,\sigma^2,r} \geq 0$, the problem is of theoretical interest as our results complement the bound in (1.1) for centered Gaussian matrices with a detailed *non-asymptotic analysis* of the noncentered case, extending also to more general rank- r projections. Finite-rank perturbations of random matrices have found recently a lot of attention, see [Capitaine et al. \(2009\)](#), [Capitaine et al. \(2012\)](#), [Pizzo et al. \(2012\)](#), [Tao \(2012\)](#) among others. [Tao \(2012\)](#), Theorem 1.7, studies the eigenvalue value spectrum of low rank perturbations of an iid complex random matrix and proves, as a special case, that $\gamma_{\max}(C + E/(\sigma\sqrt{M})) = \gamma_{\max}(C) + o_p(1)$ as $M \rightarrow \infty$ and $\operatorname{rank}(C) = O(1)$ as long as $|\gamma_{\max}(C)| = O(1)$ is sufficiently large, with $\gamma_{\max}(C)$ the eigenvalue of C which is maximal in absolute value. [Capitaine et al.](#)

(2009) and Pizzo et al. (2012) study Wigner matrices instead of iid random matrices. Somewhat remarkably, the outlier eigenvalues of the perturbed matrix are not close in probability to those of the original matrix C but to some shifted value $\lambda_i(C) + \sigma^2/\lambda_i(C)$, where σ^2 is the common variance of the entries of the Wigner matrix, and $\lambda_i(C)$ the eigenvalues of an Hermitian matrix C . Our results are complementary:

- We derive *non-asymptotic* cumulated second moment bounds on the singular values in the deformed (non-Hermitian) iid real Gaussian matrix case, i.e. the noise level σ^2 and the dimension M are fixed but arbitrary throughout the analysis, and the constants involved in our bounds do not depend on them.
- The perturbation matrix C is not required to be of low or uniformly bounded rank, for example, our results cover the case $\text{rank}(C) = \lfloor M/2 \rfloor$ or $\text{rank}(C) = M$.
- Our proofs differ significantly from the techniques of the above mentioned results but rely on empirical process techniques without making use of classical random matrix tools. The novelty in the proof of the subsequent Theorem 5.1 is that a slicing argument is used for bounding the expectation of the supremum over some *non-centered* process.

Although our results extend without difficulties to the self-adjoint dilation \tilde{X} of X in $\mathbb{R}^{2M \times 2M}$, it remains open whether, in an appropriate asymptotic sense, the eigenvalue spectrum of \tilde{X} behaves similarly to the deformed Wigner case as studied by Capitaine et al. (2009) and Pizzo et al. (2012) with finite-rank perturbations, as their assumptions do not apply to this setting.

As concerns applicability in mathematical statistics, the study of $\mathbf{E} \|\hat{\pi}_r X\|_{S_2}^2 = \sum_{i=1}^r \mathbf{E} \hat{\lambda}_i^2$ arises naturally when inferring about quantities like

$$(1.6) \quad \|C\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2 \quad \text{or} \quad \arg \min_{r \geq 1} \left\{ \frac{\|\pi_r C\|_{S_2}^2}{\|C\|_{S_2}^2} \geq \alpha \right\}, \quad \alpha \in (0, 1],$$

which are of interest for analyzing and understanding the singular value spectrum, in particular in high dimension. As above and subsequently, for any matrix $C \in \mathbb{R}^{M \times M}$, its singular values $\lambda_1, \dots, \lambda_M$ are ordered in decreasing magnitude. In terms of singular values,

$$\|C\|_{S_2}^2 = \sum_{i=1}^M \lambda_i^2 \quad \text{and} \quad \|\pi_r C\|_{S_2}^2 = \sum_{i=1}^r \lambda_i^2.$$

If $C = \sum_{i=1}^M \lambda_i U_i V_i'$ denotes some singular value decomposition of C , where U_1, \dots, U_M and V_1, \dots, V_M are two sets of orthonormal vectors in \mathbb{R}^M , then the maximizer

$$\text{Argmax}_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \mathbf{E} \|\tilde{\pi}_r(C + \mathbf{E})\|_{S_2}^2 = \text{Argmax}_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \|\tilde{\pi}_r C\|_{S_2}^2$$

is unique if and only if $\lambda_r > \lambda_{r+1}$, in which case it is equal to the orthogonal projection $\sum_{i=1}^r U_i U_i'$ onto the linear space spanned by the orthonormal column vectors U_1, \dots, U_r , and $\pi_r C = \sum_{i=1}^r \lambda_i U_i V_i'$. In the context of covariance matrices, the ratio $\|\pi_r C\|_{S_2}^2 / \|C\|_{S_2}^2$ is often referred to as percentage of the "explained variance" by the first r principal components, and the second expression in (1.6) determines the smallest number of principal components needed to explain a prescribed percentage α of the overall variance. Within our model, the statistics $\|\pi_r X\|_{S_2}^2 - \sigma^2 r M$ estimates the expression $\|\pi_r C\|_{S_2}^2$ in (1.6) unbiasedly, but note that $\pi_r = \pi_r(C)$ is not available in advance as C itself and in particular its singular spaces are unknown. Thus, the first question in the analysis is whether the empirical counterpart

$$\|\widehat{\pi}_r X\|_{S_2}^2 - \sigma^2 r M = \sum_{i=1}^r \widehat{\lambda}_i^2 - \sigma^2 r M$$

does the job as well, where $\widehat{\lambda}_1, \dots, \widehat{\lambda}_r$ denote the first r largest singular values of X . Our profound analysis about this problem will show that its bias depends strongly on the unknown matrix C itself, and even for favorable rank- r matrices $C = \mathbf{E}X$ of arbitrarily large amplitude and rank- r projections, the bias $\mathbf{E}\|\widehat{\pi}_r X\|_{S_2}^2 - \mathbf{E}\|\pi_r X\|_{S_2}^2$ remains of the order $\sigma^2 r(M - r)$, which is shown to be unimprovable in general. Note at this point that $\mathbf{E}\|\widehat{\pi}_r X\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2 \geq \mathbf{E}\|\pi_r X\|_{S_2}^2 - \|\pi_r C\|_{S_2}^2 = \sigma^2 r M$, but there is a priori no reason why the difference

$$\mathbf{E}\|\widehat{\pi}_r X\|_{S_2}^2 - \mathbf{E}\|\pi_r X\|_{S_2}^2$$

cannot be even of substantially smaller order for "good" choices C , see question (A). In order to keep the technical expenditure as small as possible, we consider the model $X = C + \mathbf{E}$ as mentioned above, but we conjecture that similar non-asymptotic implications (somewhat different and still to be derived for the Wishart case) will be valid for the squared Hilbert-Schmidt norm of rank- r projections of high-dimensional empirical covariance matrices YY' , with $Y \sim \mathcal{N}(0, C)$ for some positive semidefinite matrix C , with consequences on the robustness properties of principal component analysis.

In the simplest special cases where the question is non-trivial, the main findings of the article can be summarized as follows:

THEOREM 1.1 (Prototype of weak accuracy). *Let $C_\alpha = \alpha \text{Id}$ for some arbitrary real number $\alpha \in \mathbb{R}$, where Id denotes the $M \times M$ -identity matrix. Then the following statements hold true: (i) The case $C = C_\alpha = \alpha \text{Id}$, $\alpha \neq 0$, is always worse than the case $C = 0$:*

$$\delta_{C_\alpha, M, \sigma^2, r} \geq \delta_{0, M, \sigma^2, r}.$$

(ii) The difference $\delta_{C_\alpha, M, \sigma^2, r}$ explodes at least linearly in the amplitude $|\alpha|$: There exists some constant $c_1 > 0$, independent of σ^2 , r and M , such that

$$\liminf_{|\alpha| \rightarrow \infty} \frac{\delta_{C_\alpha, M, \sigma^2, r}}{|\alpha|} \geq c_1 \sigma r \sqrt{M - r} \quad \text{for all } r \leq M - r.$$

THEOREM 1.2 (Universal upper and lower bound in case $r = 1$). (i) *There exists some constant $c_2 > 0$ independent of C, M, σ^2 , such that for all $C \in \mathbb{R}^{M \times M}$ and $M \geq 2$*

$$(1.7) \quad \delta_{C,M,\sigma^2,1} \leq c_2 \left(\sigma^2 M + \sigma \sqrt{M} \|C\|_{S_\infty} \right),$$

where $\|C\|_{S_\infty}$ denotes the spectral norm of C .

(ii) *There exist constants $c_3 > 0$ and $M_0 \in \mathbb{N}$, independent of σ^2 , such that*

$$(1.8) \quad \inf_{C \in \mathbb{R}^{M \times M}} \delta_{C,M,\sigma^2,1} \geq c_3 \sigma^2 (M - 1) \quad \text{for all } M \geq M_0.$$

THEOREM 1.3 (Prototype of high accuracy). *Let $C_{\alpha,s} = \text{diag}(\alpha, \dots, \alpha, 0, \dots, 0)$ with $\text{rank}(C_\alpha) = s$. Then the following statements hold true: (i) *There exists some constant $c_4 > 0$ independent of α, r, M and σ^2 , such that**

$$(1.9) \quad \delta_{C_{\alpha,r},M,\sigma^2,r} \leq c_4 \sigma^2 r M \quad \text{for } r \leq M - r \text{ and every } \alpha \in \mathbb{R}.$$

(ii) *There exists some constant $c_5 > 0$ independent of M, r and σ^2 , such that the bound (1.1) is asymptotically sharp in the following sense:*

$$\liminf_{|\alpha| \rightarrow \infty} \max_{s \in \{r, M-r\}} \delta_{C_{\alpha,s},M,\sigma^2,s} \geq c_5 \sigma^2 r (M - r).$$

The same result holds true even without the maximum over $\{1, M-1\}$ for $s = r = 1$.

Theorem 1.3 (i) describes some special case of the more general upper bound in Theorem 5.1, which applies for every matrix $C \in \mathbb{R}^{M \times M}$ of $\text{rank}(C) \geq r$. In case that $\lambda_i = \alpha$ for $i \leq r$ and $\lambda_i = \beta < \alpha$ for $i > r$, the bound of Theorem 5.1 approaches the unimprovable upper bound for matrices C of constant singular value spectrum as $\beta \rightarrow \alpha$.

The article is organized as follows. In Section 2, we introduce the notation and describe some basic observations about $\delta_{C,M,\sigma^2,r}$. Prototypes of matrices C of weak accuracy and first lower bounds are studied in Section 3. The supremum over the centered differences

$$\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \left[\|\pi_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\pi_r(C + E)\|_{S_2}^2 \right] \right)$$

is analyzed in Section 4. The process of centered differences and modifications thereof are central for our analysis. Our main results are given in Section 5. Our general idea on how to derive potentially sharp upper bounds on $\delta_{C,M,\sigma^2,r}$ for general $M \times M$ -matrices C is described at the beginning of that section. The upper bounds are complemented with lower bounds in Section 6. Consequences on statistical estimation problems, in particular in the recent area of low-rank matrix recovery, are discussed in Section 7. Section 8 is devoted to the proof of Theorem 5.1. The proof of Theorem 6.2 is deferred to Section 9.

2. Preliminaries.

2.1. Notation. The notation \lesssim means less or equal up to some non-negative multiplicative constant which does not depend on the variable parameters in the expression. $A \sim B$ should be read as $A \lesssim B$ and $B \lesssim A$ at once. If not stated otherwise, \mathbf{E} is a centered Gaussian matrix whose entries are independent with variance σ^2 . Subsequently, $\|\cdot\|_{S_p}$, $1 \leq p \leq \infty$, denotes the Schatten- p -norm on $\mathbb{R}^{M \times M}$, i.e. for any $C \in \mathbb{R}^{M \times M}$, the $\|C\|_{S_p}$ coincides with the ℓ_p -norm of its singular values $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$:

$$\|C\|_{S_p} = \left(\sum_{i=1}^M \lambda_i^p \right)^{1/p} \text{ for } 1 \leq p < \infty, \text{ and } \|C\|_{S_\infty} = \lambda_1.$$

Specifically, $\|\cdot\|_{S_1}$, $\|\cdot\|_{S_2}$ and $\|\cdot\|_{S_\infty}$ are referred to as nuclear norm, Hilbert-Schmidt or Frobenius norm, and spectral norm, respectively. $\text{tr}(C)$ denotes the trace of $C \in \mathbb{R}^{M \times M}$. For any $A \in \mathbb{R}^{d \times M}$, A' denotes its transpose in $\mathbb{R}^{M \times d}$. Id denotes the $M \times M$ -identity matrix, and Id_r the diagonal matrix $\text{diag}(1, \dots, 1, 0, \dots, 0)$ of rank r . As usual, $O(M)$ describes the orthogonal group, i.e. the group of orthogonal $M \times M$ -matrices. For any totally-bounded, pseudometric space (\mathcal{X}, d) and any subset $E \subset \mathcal{X}$, the covering number $N(E, d, \delta)$ is the smallest number of closed d -balls in \mathcal{X} of radius δ needed to cover E .

2.2. Some basic observation about $\delta_{C,M,\sigma^2,r}$. The following representation clarifies the problem under consideration. For some arbitrary matrix $A \in \mathbb{R}^{M \times M}$, $\|A\|_{S_\infty} = 1$, and $\alpha \in \mathbb{R}$, inspection of the quantity $\delta_{\alpha A, \sigma^2, M, r}$ shows

$$\begin{aligned} \delta_{\alpha A, M, \sigma^2, r} &= \mathbf{E} \left\| \tilde{\pi}_r(\alpha A + \mathbf{E}) \right\|_{S_2}^2 - \mathbf{E} \left\| \pi_r(\alpha A + \mathbf{E}) \right\|_{S_2}^2 \\ &= \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\{ \left\| \tilde{\pi}_r \mathbf{E} \right\|_{S_2}^2 - \left\| \pi_r \mathbf{E} \right\|_{S_2}^2 + 2\alpha \text{tr}(\mathbf{E}'(\tilde{\pi}_r - \pi_r)A) \right. \right. \\ &\quad \left. \left. \underbrace{- \alpha^2 \left(\left\| \pi_r A \right\|_{S_2}^2 - \left\| \tilde{\pi}_r A \right\|_{S_2}^2 \right)}_{\text{"compensation term", } \leq 0} \right\} \right). \end{aligned}$$

The two processes

$$\left(\left\| \tilde{\pi}_r \mathbf{E} \right\|_{S_2}^2 - \left\| \pi_r \mathbf{E} \right\|_{S_2}^2 \right)_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \quad \text{and} \quad \left(\alpha \text{tr}(\mathbf{E}'(\tilde{\pi}_r - \pi_r)A) \right)_{\tilde{\pi}_r \in \mathcal{S}_{M,r}}$$

are centered, while the deterministic compensation term $-\alpha^2(\|\pi_r A\|_{S_2}^2 - \|\tilde{\pi}_r A\|_{S_2}^2)$ is less or equal to zero, for any choice of A and every $\tilde{\pi}_r \in \mathcal{S}_{M,r}$. Note that the stochastic term $\alpha \text{tr}(\mathbf{E}'(\tilde{\pi}_r - \pi_r)A)$ is linear in α , but the deterministic compensation term depends quadratically on α . This representation suggests that an interplay of amplitude $|\alpha|$ and structure of A determine the accuracy of approximation.

3. The prototype of weak accuracy – the case without deterministic compensation term. Recall the definition (1.5), with $\hat{\pi}_r$ and π_r as defined in (1.4). The first result is a lower bound on the expected squared Hilbert-Schmidt norm of the rank- r -projection in case that the singular value spectrum of C is constant.

PROPOSITION 3.1. *Let $C_\alpha \in \mathbb{R}^{M \times M}$ with singular value decomposition $U\Lambda_\alpha V'$. Assume that $\Lambda_\alpha = \alpha \text{Id}$ with some non-negative number $\alpha \in \mathbb{R}$. Then*

$$(i) \quad \delta_{C_\alpha, M, \sigma^2, r} \geq \delta_{0, M, \sigma^2, r} \text{ for every } \alpha > 0,$$

and

$$(ii) \quad \liminf_{\alpha \rightarrow \infty} \frac{\delta_{C_\alpha, M, \sigma^2, r}}{\alpha} \gtrsim r\sigma\sqrt{M-r} \text{ for any } r \leq M-r.$$

REMARK. Proposition 3.1 (i) demonstrates that the accuracy in case $C = \alpha \text{Id}$ is always worse than in case $C = 0$. (ii) complements this observation with an asymptotic lower bound: the difference $\delta_{C_\alpha, M, \sigma^2, r}$ explodes at least linearly in the amplitude α . In particular, (ii) provides a positive answer to question (B) in the Introduction.

PROOF. $Y =_{\mathcal{D}} Z$ for two random variables Y and Z means that their distributions coincide. Since $U'\tilde{\pi}_r U \in \mathcal{S}_{M,r}$ for any $\tilde{\pi}_r \in \mathcal{S}_{M,r}$, $U'EV =_{\mathcal{D}} E$ for two fixed orthogonal matrices U and V , and $\|U'AV\|_{S_2}^2 = \|A\|_{S_2}^2$ for any $A \in \mathbb{R}^{M \times M}$, we may assume without loss of generality that $C = \alpha \text{Id}$. Let π_r be as given in (1.4), i.e. in this case, π_r denotes some arbitrary but fixed element of $\mathcal{S}_{M,r}$. First,

$$\|\tilde{\pi}_r(\alpha \text{Id} + E)\|_{S_2}^2 - \|\pi_r(\alpha \text{Id} + E)\|_{S_2}^2 = 2\alpha \text{tr}\left(E'(\tilde{\pi}_r - \pi_r)\right) + \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2,$$

i.e. we need a lower bound on

$$\mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} 2\alpha \text{tr}\left(E'(\tilde{\pi}_r - \pi_r)\right) + \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2 \right).$$

Note that the supremum within the expectation is non-negative simply because $\pi_r \in \mathcal{S}_{M,r}$. Let $E^* \in \mathbb{R}^{M \times M}$ be the matrix with the entries $E_{ij}^* = -E_{ij}$, $i, j = 1, \dots, M$. By the symmetry of the Gaussian distribution,

$$\begin{aligned} & 2\alpha \text{tr}\left(E'(\tilde{\pi}_r - \pi_r)\right) + \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2 \\ &=_{\mathcal{D}} 2\alpha \text{tr}\left(E^{*'}(\tilde{\pi}_r - \pi_r)\right) + \|\tilde{\pi}_r E^*\|_{S_2}^2 - \|\pi_r E^*\|_{S_2}^2 \\ &= -2\alpha \text{tr}\left(E'(\tilde{\pi}_r - \pi_r)\right) + \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2. \end{aligned}$$

Consequently,

$$\begin{aligned}
& \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} 2\alpha \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) \right) + \|\tilde{\pi}_r \mathbf{E}\|_{S_2}^2 - \|\pi_r \mathbf{E}\|_{S_2}^2 \right) \\
&= \frac{1}{2} \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} 2\alpha \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) \right) + \|\tilde{\pi}_r \mathbf{E}\|_{S_2}^2 - \|\pi_r \mathbf{E}\|_{S_2}^2 \right) \\
&\quad + \frac{1}{2} \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} 2\alpha \text{tr} \left(\mathbf{E}^{*'}(\tilde{\pi}_r - \pi_r) \right) + \|\tilde{\pi}_r \mathbf{E}^*\|_{S_2}^2 - \|\pi_r \mathbf{E}^*\|_{S_2}^2 \right) \\
&= \frac{1}{2} \mathbf{E} \left[\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} 2\alpha \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) \right) + \|\tilde{\pi}_r \mathbf{E}\|_{S_2}^2 - \|\pi_r \mathbf{E}\|_{S_2}^2 \right. \\
&\quad \left. + \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} 2\alpha \text{tr} \left(\mathbf{E}^{*'}(\tilde{\pi}_r - \pi_r) \right) + \|\tilde{\pi}_r \mathbf{E}^*\|_{S_2}^2 - \|\pi_r \mathbf{E}^*\|_{S_2}^2 \right] \\
&\geq \frac{1}{2} \mathbf{E} \left[\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left(2\alpha \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) \right) + \|\tilde{\pi}_r \mathbf{E}\|_{S_2}^2 - \|\pi_r \mathbf{E}\|_{S_2}^2 \right. \right. \\
&\quad \left. \left. + 2\alpha \text{tr} \left(\mathbf{E}^{*'}(\tilde{\pi}_r - \pi_r) \right) + \|\tilde{\pi}_r \mathbf{E}^*\|_{S_2}^2 - \|\pi_r \mathbf{E}^*\|_{S_2}^2 \right) \right] \\
&= \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \|\tilde{\pi}_r \mathbf{E}\|_{S_2}^2 - \|\pi_r \mathbf{E}\|_{S_2}^2 \right),
\end{aligned}$$

which proves part (i) of the proposition. As concerns the proof of (ii), observe that

$$\begin{aligned}
& \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} 2\alpha \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) \right) + \|\tilde{\pi}_r \mathbf{E}\|_{S_2}^2 - \|\pi_r \mathbf{E}\|_{S_2}^2 \right) \\
&\geq 2\alpha \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) \right) \right) - \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \|\pi_r \mathbf{E}\|_{S_2}^2 - \|\tilde{\pi}_r \mathbf{E}\|_{S_2}^2 \right) \\
(3.1) \quad &\geq 2\alpha \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) \right) \right) - \sigma^2 r M,
\end{aligned}$$

since $\mathbf{E}(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \|\pi_r \mathbf{E}\|_{S_2}^2 - \|\tilde{\pi}_r \mathbf{E}\|_{S_2}^2) \leq \sigma^2 r M$. By Sudakov's minoration, there exists some universal constant c_{Sud} such that

$$(3.2) \quad \mathbf{E} \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) \right) \geq \sigma c_{\text{Sud}} \delta \sqrt{\log N(\mathcal{S}_{M,r}, d_{S_2}, \delta)}$$

for any $\delta > 0$. Proposition 8 in [Pajor \(1998\)](#) states for any $r \leq M - r$

$$\left(\frac{c'}{\xi} \right)^{r(M-r)} \leq N(\mathcal{S}_{M,r}, d_{S_2}, \xi \sqrt{r}), \quad \forall \xi > 0,$$

with some universal constant $c' > 0$ which does not depend on r and M . Choosing $\delta = \sqrt{r}c'/e$ and plugging (3.2) into (3.1) yields, for some constant c which does not depend on σ^2, r, M and α :

$$\delta_{\alpha \text{Id}, M, \sigma^2, r} \geq c \sigma r \alpha \sqrt{M - r} - \sigma^2 r M.$$

Dividing both sides by α and taking the limes inferior proves (ii). \square

In view of the representation in Section 2, the case $C = \alpha \text{Id}$ is the prototype of weak accuracy as there is no deterministic compensation term in the expression of the supremum. It follows from the subsequent Corollary 4.2 that the lower bound of Proposition 3.1 (ii) is sharp.

4. S_2 - S_∞ -chaining bounds for the supremum over the centered process and first consequences on $\delta_{C,M,\sigma^2,r}$. Let $X = C + E$ as described in (1.3). Recall the definition from the Introduction

$$\pi_r \in \underset{\tilde{\pi}_r \in \mathcal{S}_{M,r}}{\text{Argmax}} \mathbf{E} \|\tilde{\pi}_r X\|_{S_2}^2 = \underset{\tilde{\pi}_r \in \mathcal{S}_{M,r}}{\text{Argmax}} \|\tilde{\pi}_r C\|_{S_2}^2.$$

Because of $\mathbf{E} \|\pi_r(C + E)\|_{S_2}^2 \geq \mathbf{E} \|\tilde{\pi}_r(C + E)\|_{S_2}^2$ for every $\tilde{\pi}_r \in \mathcal{S}_{M,r}$, it follows that

$$\begin{aligned} (4.1) \quad & \mathbf{E} \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \|\pi_r(C + E)\|_{S_2}^2 \right) \\ & \leq \mathbf{E} \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\tilde{\pi}_r(C + E)\|_{S_2}^2 \right. \\ & \quad \left. - \left[\|\pi_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\pi_r(C + E)\|_{S_2}^2 \right] \right). \end{aligned}$$

That is, the study of the supremum over the centered process

$$Z := \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \left[\|\pi_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\pi_r(C + E)\|_{S_2}^2 \right] \right).$$

yields some first (possibly very rough) estimate of $\delta_{C,M,\sigma^2,r}$ from above. Variants thereof are central for our subsequent analysis in Section 5.

PROPOSITION 4.1. *Let $(E_{ij})_{i,j=1}^M$ be a centered matrix of iid Gaussian entries with variance σ^2 . Then there exists some constant $c > 0$ such that for every $1 \leq r < M \in \mathbb{N}$ and every $C \in \mathbb{R}^{M \times M}$*

$$(4.2) \quad \mathbf{E} Z \leq c \left(\sigma^2 r M + \sigma r \sqrt{M} \|C\|_{S_\infty} \right).$$

The proof of Proposition 4.1 is deferred to the end of this section. We draw some first consequences on $\delta_{C,M,\sigma^2,r}$.

COROLLARY 4.2. (i) (Universal upper bound) For all $C \in \mathbb{R}^{M \times M}$,

$$\delta_{C,M,\sigma^2,r} \lesssim \sigma^2 r M + \sigma r \sqrt{M} \|C\|_{S_\infty}.$$

(ii) (Upper bound in the "small amplitude" regime)

$$\sup_{\substack{C \in \mathbb{R}^{M \times M}, \\ \|C\|_{S_\infty} \leq \sigma \sqrt{M}}} \delta_{C,M,\sigma^2,r} \lesssim \sigma^2 r M.$$

PROOF of the Corollary 4.2 (i) follows from (4.1) and the bound on $\mathbf{E} Z$ given in Proposition 4.1; (ii) follows from (i).

REMARK. It is worth being mentioned that $\mathbf{E} Z$ grows linearly with $\|C\|_{S_\infty}$, and this linear dependence is optimal for matrices $X = \alpha \text{Id} + E$, $\alpha \in \mathbb{R}$, in view of Proposition 3.1. In particular, the lower bound of Proposition 3.1 (ii) is sharp, and the universal upper bound cannot be improved without further structural assumptions on C .

The next lemma complements the upper bound of the Corollary 4.2 (ii) in the small amplitude regime with a lower bound.

LEMMA 4.3. *For any real constant $\kappa > 0$,*

$$\inf_{\substack{C \in \mathbb{R}^{M \times M}; \\ \|C\|_{S_\infty} \leq \kappa \sigma \sqrt{M}}} \delta_{C,M,\sigma^2,r} \geq \delta_{0,M,\sigma^2,r} - \kappa^2 \sigma^2 r M.$$

PROOF. With the same symmetry argument as used in the proof of Proposition 3.1 (i), we obtain for any $C \in \mathbb{R}^{M \times M}$ with $\|C\|_{S_\infty} \leq \kappa \sigma \sqrt{M}$

$$\begin{aligned} & \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\| \tilde{\pi}_r(C + E) \right\|_{S_2}^2 - \left\| \pi_r(C + E) \right\|_{S_2}^2 \right) \\ & \geq \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\| \tilde{\pi}_r E \right\|_{S_2}^2 - \left\| \pi_r E \right\|_{S_2}^2 + \left\| \tilde{\pi}_r C \right\|_{S_2}^2 - \left\| \pi_r C \right\|_{S_2}^2 \right) \\ & \geq \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\| \tilde{\pi}_r E \right\|_{S_2}^2 - \left\| \pi_r E \right\|_{S_2}^2 \right) - \kappa^2 \sigma^2 r M. \quad \square \end{aligned}$$

REMARK. Together with the upper Bai-Yin bound in expectation (1.1), Lemma 4.3 implies in particular for $r = 1$ that there exists some $M_0 \in \mathbb{N}$, independent of σ^2 , such that

$$\inf_{\substack{C \in \mathbb{R}^{M \times M}; \\ \|C\|_{S_\infty} \leq \sigma \sqrt{M}}} \delta_{C,M,\sigma^2,1} \geq \frac{1}{2} \delta_{0,M,\sigma^2,1} \sim \sigma^2 M \quad \text{for all } M \geq M_0.$$

That is, in the small amplitude regime $\|C\|_{S_\infty} \leq \sigma \sqrt{M}$, the accuracy is never better in order than in the case $C = 0$, independent of the specific structure of the matrix C .

PROOF OF PROPOSITION 4.1. For any r -dimensional subspace $U \subset \mathbb{R}^M$, let $P_U \in \mathcal{S}_{M,r}$ denote the orthogonal projection onto U . The proof is based on the classical generic chaining device. In order to make this technique applicable, we need to investigate pairwise differences of the centered process $Z^{\sigma,M,r}$ which is

pointwise given by

$$\begin{aligned} Z_{P_U}^{\sigma, M, r} &:= \left(\|P_U(C + E)\|_{S_2}^2 - \mathbf{E} \|P_U(C + E)\|_{S_2}^2 \right) \\ &= \left(\text{tr}(E' P_U E) - r M \sigma^2 + 2 \text{tr}(C' P_U E) \right), \quad P_U \in \mathcal{S}_{M, r}. \end{aligned}$$

Denote $\pi_r^{(1)} = P_{U_1} \in \mathcal{S}_{M, r}$, $\pi_r^{(2)} = P_{U_2} \in \mathcal{S}_{M, r}$, and $A = \pi_r^{(1)} - \pi_r^{(2)}$. Recall that $P'_{U_i} = P_{U_i}$ and $P_{U_i}^2 = P_{U_i}$ for $i = 1, 2$. For any $B = (b_{ij})_{i,j=1}^M \in \mathbb{R}^{M \times M}$, $\text{vec}(B)$ denotes the associated vector obtained by sticking together its columns,

$$\text{vec}(B) := (b_{11}, \dots, b_{M1}, b_{12}, \dots, b_{M2}, \dots, b_{MM})' \in \mathbb{R}^{M^2}.$$

Observe furthermore that

$$\begin{aligned} \|P_{U_1} E\|_{S_2}^2 - \|P_{U_2} E\|_{S_2}^2 &= \text{tr}(E' (P'_{U_1} P_{U_1} - P'_{U_2} P_{U_2}) E) \\ &= \sum_{l,k,i=1}^M A_{lk} E_{ki} E_{li} = \text{vec}(E)' \tilde{A} \text{vec}(E), \end{aligned}$$

where \tilde{A} denotes the block-diagonal matrix $\text{diag}(A, \dots, A) \in \mathbb{R}^{M^2 \times M^2}$, and, analogously,

$$\text{tr}(C' P_{U_1} E) - \text{tr}(C' P_{U_2} E) = \text{vec}(C)' \tilde{A} \text{vec}(E).$$

Noting that $\|\tilde{A}\|_{S_2} = \sqrt{M} \|A\|_{S_2}$ and $\|\tilde{A}\|_{S_\infty} = \|A\|_{S_\infty}$, Bernstein's inequality for quadratic forms of Gaussian variables (see, for instance, [Bechar \(2009\)](#), Lemma 0.2) yields the exponential bound

$$\begin{aligned} (4.3) \quad &\mathbb{P} \left(Z_{P_{U_1}}^{\sigma, M, r} - Z_{P_{U_2}}^{\sigma, M, r} \geq \right. \\ &\left. 2\sqrt{\sigma^4 M \|P_{U_1} - P_{U_2}\|_{S_2}^2 + 2\sigma^2 \|\tilde{A} \text{vec}(C)\|_2^2} \cdot \sqrt{t} + 2\sigma^2 \|P_{U_1} - P_{U_2}\|_{S_\infty} t \right) \leq \exp(-t) \end{aligned}$$

for all $t > 0$. Note that the bound is fully symmetric in U_1 and U_2 . Since $\|\tilde{A} \text{vec}(C)\|_2^2 = \|AC\|_{S_2}^2 \leq \|A'A\|_{S_1} \|C'C\|_{S_\infty} = \|A\|_{S_2}^2 \|C\|_{S_\infty}^2$, it follows that

$$(4.4) \quad \sqrt{\sigma^4 M \|P_{U_1} - P_{U_2}\|_{S_2}^2 + 2\sigma^2 \|\tilde{A} \text{vec}(C)\|_2^2} \leq \sqrt{\sigma^4 M + 2\sigma^2 \|C\|_{S_\infty}^2} \|P_{U_1} - P_{U_2}\|_{S_2},$$

i.e. the exponential tail bound for the differences $Z_{P_{U_1}}^{\sigma, M, r} - Z_{P_{U_2}}^{\sigma, M, r}$ is characterized by an interplay of Hilbert-Schmidt and spectral norm, which take over the roles of the ℓ_2 - and ℓ_∞ -norms of the classical Bernstein inequality from the vector case. We prove first the case $r \leq M - r$. Note that

$$(4.5) \quad \sup_{\pi_r, \tilde{\pi}_r \in \mathcal{S}_{M, r}} \|\pi_r - \tilde{\pi}_r\|_{S_2}^2 \leq 2r \quad \text{and} \quad \sup_{\pi_r, \tilde{\pi}_r \in \mathcal{S}_{M, r}} \|\pi_r - \tilde{\pi}_r\|_{S_\infty} \leq 2.$$

Since $Z_x^{\sigma, M, r}$ depends continuously on x , it holds that

$$\mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M, r}} (Z_{\tilde{\pi}_r}^{\sigma, M, r} - Z_{\pi_r}^{\sigma, M, r}) \right) = \mathbf{E} \left(\sup_{\tilde{\pi}_r \in S} (Z_{\tilde{\pi}_r}^{\sigma, M, r} - Z_{\pi_r}^{\sigma, M, r}) \right)$$

for any countable, dense subset S of $\mathcal{S}_{M, r}$, and by the Theorem of monotone convergence, it is sufficient to assume subsequently that S is finite. We define now recursively an increasing family of partitions $(\mathcal{A}_n)_{n \geq 0}$ of S such that $\mathcal{A}_0 = S$, and for $n \geq 1$ and $A \in \mathcal{A}_n$

$$(4.6) \quad (i) \quad \|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_2} \leq 2^{-n} \sqrt{2r} \quad \text{and} \quad (ii) \quad \|\pi_r^{(1)} - \pi_r^{(2)}\|_{S_\infty} \leq 2^{-n+1}$$

$\forall \pi_r^{(1)}, \pi_r^{(2)} \in A$, with $\mathcal{A}_{n+1} \subset \mathcal{A}_n$ for all $n \geq 0$. This can be realized as follows: For $n = 0$, $\mathcal{A}_{0,2} := \mathcal{A}_{0,\infty} := \{S\}$. For any totally-bounded, pseudometric space (\mathcal{X}, d) and any subset $E \subset \mathcal{X}$, the covering number $N(E, d, \delta)$ is the smallest number of closed d -balls in \mathcal{X} of radius δ needed to cover E . It is proved in [Szarek \(1982\)](#), see also [Pajor \(1998\)](#) for a different proof, that for $1 \leq r \leq M - r$ and any $C' > \xi > 0$

$$(4.7) \quad N(\mathcal{S}_{M, r}, d_{S_2}, \xi \sqrt{r}) \leq \left(\frac{C'}{\xi} \right)^{r(M-r)} \quad \text{and} \quad N(\mathcal{S}_{M, r}, d_{S_\infty}, \xi) \leq \left(\frac{C'}{\xi} \right)^{r(M-r)}$$

for some universal constant $C' > 0$. Hence, S can be covered with at most $(C')^{r(M-r)}$ S_2 -balls $B_{1,2}, \dots, B_{N_{1,2},2}$ of radius $\sqrt{2r}$, and with at most $(C')^{r(M-r)}$ S_∞ -balls $B_{1,\infty}, \dots, B_{N_{1,\infty},\infty}$ of radius 2. From such finite coverings of S_2 - and S_∞ -balls, the partitions $\mathcal{A}_{1,2}$ and $\mathcal{A}_{1,\infty}$ are canonically constructed by

$$\mathcal{A}_{1,j} := \left\{ \left(B_{k,j} \setminus \bigcup_{1 \leq l < k} B_{l,j} \right) \cap S, k = 1, \dots, N_{1,j} \right\}, \quad j = 2, \infty.$$

For $n \geq 2$ we proceed inductively using the bounds (8.11) and (8.15) for $q = \infty$ of Lemma 8.2. Indeed, each element $A \in \mathcal{A}_{n-1,j}$ is element of an S_j -ball in $\mathcal{S}_{M, r}$ of radius $2^{-n} \sqrt{2r}$ and 2^{-n+1} , respectively, and can be partitioned as above into $(2C')^{r(M-r)}$ subsets of balls of radius $2^{-(n+1)} \sqrt{2r}$ and 2^{-n} , respectively. By construction, the partitions $(\mathcal{A}_{n,2})_{n \geq 0}$ and $(\mathcal{A}_{n,\infty})_{n \geq 0}$ are nested, and $\text{card}(\mathcal{A}_{n,2}) \leq D^{nr(M-r)}$, $\text{card}(\mathcal{A}_{n,\infty}) \leq D^{nr(M-r)}$ for some universal constant $D > 0$. Setting now $\mathcal{A}_n := \{A_2 \cap A_\infty : A_2 \in \mathcal{A}_{n,2} \text{ and } A_\infty \in \mathcal{A}_{n,\infty}\}$ yields some partition with the above mentioned properties (4.6). Obviously,

$$(4.8) \quad \text{card}(\mathcal{A}_n) \leq \text{card}(\mathcal{A}_{n,2}) \text{card}(\mathcal{A}_{n,\infty}).$$

For each $n \geq 1$ and $A \in \mathcal{A}_n$, let $s_n(A)$ be some arbitrarily chosen rank- r -projection matrix in A . For each $s \in \mathcal{S}$ and $n \geq 1$, there exists some unique $A \in \mathcal{A}_n$ with $s \in A$, and we set $\Pi_n(s) = s_n(A)$. When $n = 0$, define $\Pi_0(s) = \pi_r$. Now,

$$(4.9) \quad \mathbf{E} \left[\sup_{s \in \mathcal{S}} \left(Z_s^{\sigma, M, r} - Z_{\pi_r}^{\sigma, M, r} \right) \right] = \mathbf{E} \left[\sup_{s \in \mathcal{S}} \sum_{n \geq 0} \left(Z_{\Pi_{n+1}(s)}^{\sigma, M, r} - Z_{\Pi_n(s)}^{\sigma, M, r} \right) \right]$$

$$(4.10) \quad \leq \sum_{n \geq 0} \mathbf{E} \left[\sup_{s \in \mathcal{S}} \left(Z_{\Pi_{n+1}(s)}^{\sigma, M, r} - Z_{\Pi_n(s)}^{\sigma, M, r} \right) \right].$$

Note that the decomposition $\sum_{n \geq 0} Z_{\Pi_{n+1}(s)}^{\sigma, M, r} - Z_{\Pi_n(s)}^{\sigma, M, r}$ of each $Z_s^{\sigma, M, r} - Z_{\pi_r}^{\sigma, M, r}$ in (4.9) is finite since \mathcal{S} is finite. By construction,

$$\sup_{s \in \mathcal{S}} \|\Pi_n(s) - s\|_{S_2} \leq 2^{-n} \sqrt{2r} \quad \text{and} \quad \sup_{s \in \mathcal{S}} \|\Pi_n(s) - s\|_{S_\infty} \leq 2^{-n+1},$$

hence $\|\Pi_{n+1}(s) - \Pi_n(s)\|_{S_2} \leq 3 \cdot 2^{-n} \sqrt{2r}$ and $\|\Pi_{n+1}(s) - \Pi_n(s)\|_{S_\infty} \leq 3 \cdot 2^{-n+1}$, for every $s \in \mathcal{S}$. Note that $\text{card}(\{\Pi_{n+1}(s) - \Pi_n(s) : s \in \mathcal{S}\}) \leq \text{card}(\mathcal{A}_{n+1}) \text{card}(\mathcal{A}_n)$, $n \in \mathbb{N}$. Applying now Lemma A.1, [van der Vaart \(1996\)](#), to each expectation within (4.10), yields

$$\begin{aligned} & \mathbf{E} \left[\sup_{s \in \mathcal{S}} \left(Z_s^{\sigma, M, r} - Z_{\pi_r}^{\sigma, M, r} \right) \right] \\ & \lesssim \sum_{n \geq 0} \frac{1}{2^n} \left(\sqrt{2r} \sqrt{\sigma^4 M + 2\sigma^2 \|C\|_{S_\infty}^2} \sqrt{2 \log(N_n)} + 2\sigma^2 \log(N_n) \right) \\ & \lesssim \sigma^2 r M + \sigma r \sqrt{M} \|C\|_{S_\infty}, \end{aligned}$$

where $N_n := \text{card}(\mathcal{A}_{n+1}) \text{card}(\mathcal{A}_n) \leq D^{(4n+2)r(M-r)}$. Note further that the case $r = M$ is obvious, the case $M > r > M - r$ follows by consideration of the orthogonal complements $\|P_U(C + E)\|_{S_2}^2 = \|C + E\|_{S_2}^2 - \|P_U^\perp(C + E)\|_{S_2}^2$. \square

5. The main result – a general expectation bound for non-centered Gaussian matrices. Corollary 4.2 (i) provides some upper bound on $\delta_{C, M, \sigma^2, r}$ which is valid for every $M \times M$ -matrix C , and which is achieved, for instance, for $C = 0$ and $r = 1$ in the small amplitude regime, and for $C = \alpha \text{Id}$ with $|\alpha|$ sufficiently large in the large amplitude regime, i.e. $|\alpha| \gg \sigma \sqrt{M}$. In this section, we present some new and more refined analysis for bounding $\delta_{C, M, \sigma^2, r}$ which takes advantage of some potentially favorable structure of C - resulting in the presence of the deterministic compensation term as explained in Section 2. Some prototype of matrices of "high accuracy" in the large amplitude regime is discovered and analyzed.

Our approach is motivated and explained in what follows. The conjecture about the possibility of improvement over Corollary 4.2 (i) for a certain type of matrices follows from the fact that, in contrast to the situation in Section 3, the differences

$$(5.1) \quad \|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \|\pi_r(C + E)\|_{S_2}^2$$

are usually not centered. With $\pi_r = \pi_r(C)$ maximizing the expression $\|\tilde{\pi}_r C\|_{S_2}^2$ over $\mathcal{S}_{M, r}$ here and subsequently, the expectation of (5.1) is less or equal to zero. Depending on the structure of C , it may be substantially smaller than zero for some

appreciable amount of rank- r -projections. Consequently, for any subset $\mathcal{A}_C \subset \mathcal{S}_{M,r}$,

$$(5.2) \quad \begin{aligned} & \sup_{\tilde{\pi}_r \in \mathcal{A}_C} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \|\pi_r(C + E)\|_{S_2}^2 \right) \\ & \leq \sup_{\tilde{\pi}_r \in \mathcal{A}_C} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\tilde{\pi}_r(C + E)\|_{S_2}^2 \right. \\ & \quad \left. - \left[\|\pi_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\pi_r(C + E)\|_{S_2}^2 \right] \right) - \Delta_{\mathcal{A}_C}, \end{aligned}$$

$$\text{with } \Delta_{\mathcal{A}_C} := \inf_{\tilde{\pi}_r \in \mathcal{A}_C} \left(\mathbf{E} \|\pi_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\tilde{\pi}_r(C + E)\|_{S_2}^2 \right).$$

Roughly speaking, the expectation of the supremum in (5.2) is small if the supremum of the corresponding centered process over \mathcal{A}_C is small as compared to $\Delta_{\mathcal{A}_C}$. The idea for the general bound is based on decomposing the Grassmann manifold along the geometric grid of slices $\mathcal{A}_{C,k}$, $k \in \mathbb{N}$:

$$\mathcal{A}_{C,k} := \left\{ \tilde{\pi}_r \in \mathcal{S}_{M,r} : \frac{\|\pi_r C\|_{S_2}^2}{2^{k+1}} < \mathbf{E} \|\pi_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\tilde{\pi}_r(C + E)\|_{S_2}^2 \leq \frac{\|\pi_r C\|_{S_2}^2}{2^k} \right\}.$$

$$\text{Define the random variables } Y_k := \sup_{\tilde{\pi}_r \in \mathcal{A}_{C,k}} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \|\pi_r(C + E)\|_{S_2}^2 \right)$$

and

$$Y_k^0 = \sup_{\tilde{\pi}_r \in \mathcal{A}_{C,k}} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \left[\|\pi_r(C + E)\|_{S_2}^2 - \mathbf{E} \|\pi_r(C + E)\|_{S_2}^2 \right] \right).$$

$$\text{With } \tilde{Z} := \sup_{\substack{\tilde{\pi}_r \in \mathcal{S}_{M,r} : \\ \|\tilde{\pi}_r C\|_{S_2} = \|\pi_r C\|_{S_2}}} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \|\pi_r(C + E)\|_{S_2}^2 \right),$$

we obtain the series expansion

$$(5.3) \quad \begin{aligned} \delta_{C, M\sigma^2, r} &= \mathbf{E} \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left(\|\tilde{\pi}_r(C + E)\|_{S_2}^2 - \|\pi_r(C + E)\|_{S_2}^2 \right) \\ &\leq \mathbf{E} \tilde{Z} + \sum_{k \in \mathbb{N}} \mathbf{E} (0 \vee Y_k) \\ &\leq \mathbf{E} \tilde{Z} + \sum_{k \in \mathbb{N}} \mathbf{E} (0 \vee (Y_k^0 - \Delta_{\mathcal{A}_{C,k}})). \end{aligned}$$

In view of this expansion, it is clear that good bounds can be reached in principal only if $\mathbf{E} Y_k^0$ is small as compared to $\Delta_{\mathcal{A}_{C,k}}$. The evaluation of the expectations

$$\mathbf{E} (0 \vee (Y_k^0 - \Delta_{\mathcal{A}_{C,k}})), \quad k \in \mathbb{N},$$

is however quite a difficult task in general: It requires some suitable characterization of the subsets $\mathcal{A}_{C,k}$ of the Grassmannian for general matrices $C \in \mathbb{R}^{M \times M}$ and tight bounds on their metric entropy. Before discussing this serious issue, we present first the main result of this section.

THEOREM 5.1. *Let $(E_{ij})_{i,j=1}^M$ be a centered matrix of independent Gaussian entries with variance σ^2 . Then for any $C \in \mathbb{R}^{M \times M}$ with $\text{rank}(C) \geq r$ and $r \leq M - r$, the following bound holds true:*

$$\delta_{C,M,\sigma^2,r} \lesssim \sigma^2 r M \left(\text{I} + \min(\text{II}, \text{III}) \right),$$

where, with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ denoting the singular values of C ,

$$\begin{aligned} \text{I} &= \min \left(\frac{\lambda_1^2}{\lambda_r^2}, 1 + \frac{\lambda_1}{\sigma\sqrt{M}} \right), \\ \text{II} &= \left(\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2} \right)^{1/2} \cdot \frac{\lambda_1}{\sigma\sqrt{M}}, \quad \text{and} \\ \text{III} &= \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2} \quad \text{if } \lambda_{r+1} < \lambda_r, \quad \text{and } \text{III} = \infty \quad \text{else.} \end{aligned}$$

In particular for the case $r = 1$, the bound applies to any $0 \neq C \in \mathbb{R}^{M \times M}$. Some immediate consequence of Theorem 5.1 is the following.

COROLLARY 5.2. *Let $C_{\alpha,\beta,r} \in \mathbb{R}^{M \times M}$ with singular values $\lambda_i = \alpha$ for $i \leq r$ and $\lambda_i = \beta \leq \alpha$ for $i > r$. As usual, set $c/0 := \infty$ for any $c > 0$. Then*

$$\delta_{C_{\alpha,\beta,r},M,\sigma^2,r} \lesssim \sigma^2 r M \left\{ 1 + \min \left(\frac{\alpha^2}{\alpha^2 - \beta^2}, \frac{\beta}{\sigma\sqrt{M}} \right) \right\} \quad \text{for every } 0 \leq \beta \leq \alpha.$$

REMARK. Corollary (5.2) covers the two extreme cases:

- (i) Prototype of high accuracy in the "large amplitude"-regime $\alpha \gg \sigma\sqrt{M}$:

$$\delta_{C_{\alpha,0,r},M,\sigma^2,r} \lesssim \sigma^2 r M \quad \text{for every } \alpha \geq 0.$$

- (ii) Prototype of weak accuracy in the "large amplitude"-regime $\alpha \gg \sigma\sqrt{M}$: the bound of Corollary 5.2 approaches the unimprovable upper bound for matrices C of constant singular value spectrum as $\beta \rightarrow \alpha$, see Proposition 3.1.

Note that the upper bound in case $C = 0$ is covered by Proposition 4.1.

The proof of Theorem 5.1 is deferred to Section 8. Subsection 8.1 deals with the description of the sets $\mathcal{G}_{M,r}(\delta) := \{\tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \leq \delta\}$, which characterize the slices $\mathcal{A}_{C,k} = \mathcal{G}_{M,r}(2^{-k} \|\pi_r C\|_{S_2}^2) \setminus \mathcal{G}_{M,r}(2^{-(k+1)} \|\pi_r C\|_{S_2}^2)$. It is shown that these sets can be approximated by sets of very simple geometric structure. In case of some substantial spectral gap in r , this approximation is very tight. Sharp bounds on their metric entropy are derived in Subsection 8.2. The final arguments differ slightly from the description at the beginning of this Section. They are given in Subsection 8.3.

6. Lower bounds. The best possible upper bound for the accuracy of approximation in Theorem 5.1 is of the order $\sigma^2 r M$, which is attained for matrices $C = 0$ or $C = \alpha \text{Id}_r$ in the large amplitude regime. The question arises whether this bound is sharp, i.e. whether it indicates some fundamental limit on the accuracy of approximation. For fixed $A \in \mathbb{R}^{M \times M}$ with $\|A\|_{S_\infty} = 1$ and some arbitrary real number $\alpha \in \mathbb{R}$, inspection of $\delta_{\alpha A, M, \sigma^2, r}$ shows

$$(6.1) \quad \begin{aligned} & \mathbf{E} \left\| \hat{\pi}_r(\alpha A + \mathbf{E}) \right\|_{S_2}^2 - \mathbf{E} \left\| \pi_r(\alpha A + \mathbf{E}) \right\|_{S_2}^2 \\ &= \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M, r}} \left\| \tilde{\pi}_r \mathbf{E} \right\|_{S_2}^2 - \left\| \pi_r \mathbf{E} \right\|_{S_2}^2 + 2\alpha \text{tr} \left(\mathbf{E}' (\tilde{\pi}_r - \pi_r) A \right) \right. \end{aligned}$$

$$(6.2) \quad \left. - \alpha^2 \left(\left\| \pi_r A \right\|_{S_2}^2 - \left\| \tilde{\pi}_r A \right\|_{S_2}^2 \right) \right).$$

Now, $\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \geq 0$ since π_r optimizes the Hilbert-Schmidt norm, and the dependence on α in the deterministic compensation term in line (6.2) is quadratic while it is only linear in the stochastic part in line (6.1). So, for any fixed σ^2 , r , M and A , one may wonder whether the accuracy of approximation $\mathbf{E} \|\hat{\pi}_r(\alpha \text{Id}_r + \mathbf{E})\|_{S_2}^2 - \mathbf{E} \|\pi_r(\alpha \text{Id}_r + \mathbf{E})\|_{S_2}^2$ tends even to zero as $|\alpha|$ goes to infinity if A is suitably chosen, like, for instance, $A = \text{Id}_r$. In this section, we demonstrate that this is not the case. We provide some complete proof of the conjecture

$$(6.3) \quad \inf_{C \in \mathbb{R}^{M \times M}} \delta_{C, M, \sigma^2, r} \gtrsim \delta_{0, M, \sigma^2, r}$$

in case $r = 1$. For $r > 1$ we present some partial solution in the large amplitude regime.

6.1. The universal lower bound for $r = 1$.

THEOREM 6.1. *Let $(\mathbf{E}_{ij})_{i,j=1}^M$ be a centered matrix of independent Gaussian entries with variance σ^2 . Then there exists some $M_0 \in \mathbb{N}$, independent of σ^2 , such that*

$$(6.4) \quad \inf_{C \in \mathbb{R}^{M \times M}} \delta_{C, M, \sigma^2, 1} \gtrsim \delta_{0, M, \sigma^2, 1}$$

for any $M \geq M_0$.

PROOF. In view of Lemma 4.3 and its subsequent remark, it is sufficient to prove that for every $\beta > 0$, there exists some constant $c_\beta > 0$, independent of σ^2 and M , such that

$$(6.5) \quad \inf_{\substack{C \in \mathbb{R}^{M \times M}, \\ \|C\|_{S_\infty} \geq \beta \sigma \sqrt{M}}} \geq c_\beta \sigma^2 (M - 1) \quad \text{for all } M \geq 2.$$

Let $C = \sum_{i=1}^M \lambda_i U_i V_i'$ denote some singular value decomposition of C , and define $\pi_1 := U_1 U_1'$. Since for any $1 \leq s < M$ and $\pi_s, \tilde{\pi}_s \in \mathcal{S}_{M, s}$, $\pi_s - \tilde{\pi}_s = (\text{Id} - \pi_s) \tilde{\pi}_s -$

$\pi_s(\text{Id} - \tilde{\pi}_s)$ is an orthogonal decomposition and $\|\pi_s(\text{Id} - \tilde{\pi}_s)\|_{S_2}^2 = \|(\text{Id} - \pi_s)\tilde{\pi}_s\|_{S_2}^2$, observe that

$$\begin{aligned} \|\pi_1 C\|_{S_2}^2 - \|\tilde{\pi}_1 C\|_{S_2}^2 &\leq \|\pi_1 C\|_{S_2}^2 - \|\tilde{\pi}_1 \pi_1 C\|_{S_2}^2 \\ &= \lambda_1^2 \left(\|\pi_1\|_{S_2}^2 - \|\tilde{\pi}_1 \pi_1\|_{S_2}^2 \right) \\ &= \lambda_1^2 \|(\text{Id} - \tilde{\pi}_1) \pi_1\|_{S_2}^2 \\ &= \lambda_1^2 \|\tilde{\pi}_1(\text{Id} - \pi_1)\|_{S_2}^2 = \frac{\lambda_1^2}{2} \|\tilde{\pi}_1 - \pi_1\|_{S_2}^2. \end{aligned}$$

Consequently,

$$\begin{aligned} \delta_{C,M,\sigma^2,1} &= \mathbf{E} \left\| \hat{\pi}_1(C + \mathbf{E}) \right\|_{S_2}^2 - \mathbf{E} \left\| \pi_1(C + \mathbf{E}) \right\|_{S_2}^2 \\ &= \mathbf{E} \left(\sup_{\tilde{\pi}_1 \in \mathcal{S}_{M,r}} \left\{ \|\tilde{\pi}_1 \mathbf{E}\|_{S_2}^2 - \|\pi_1 \mathbf{E}\|_{S_2}^2 + 2 \text{tr}(\mathbf{E}'(\tilde{\pi}_1 - \pi_1)C) \right. \right. \\ &\quad \left. \left. - \left(\|\pi_1 C\|_{S_2}^2 - \|\tilde{\pi}_1 C\|_{S_2}^2 \right) \right\} \right) \\ &\geq \mathbf{E} \left(\sup_{\tilde{\pi}_1 \in \mathcal{S}_{M,r}} \left\{ \|\tilde{\pi}_1 \mathbf{E}\|_{S_2}^2 - \|\pi_1 \mathbf{E}\|_{S_2}^2 + 2 \text{tr}(\mathbf{E}'(\tilde{\pi}_1 - \pi_1)C) \right. \right. \\ &\quad \left. \left. - \frac{\lambda_1^2}{2} \|\pi_1 - \tilde{\pi}_1\|_{S_2}^2 \right\} \right) \\ (6.6) \quad &\geq \mathbf{E} \left(\sup_{\tilde{\pi}_1 \in \mathcal{S}_{M,1}(\delta)} \left\{ \|\tilde{\pi}_1 \mathbf{E}\|_{S_2}^2 - \|\pi_1 \mathbf{E}\|_{S_2}^2 + 2 \text{tr}(\mathbf{E}'(\tilde{\pi}_1 - \pi_1)C) \right. \right. \\ (6.7) \quad &\quad \left. \left. - \frac{\lambda_1^2}{2} \|\pi_1 - \tilde{\pi}_1\|_{S_2}^2 \right\} \right) \end{aligned}$$

for any subset

$$\mathcal{S}_{M,1}(\delta) := \left\{ \tilde{\pi}_1 \in \mathcal{S}_{M,1} : \|\tilde{\pi}_1 - \pi_1\|_{S_2} \leq \sqrt{2\delta} \right\}, \quad \delta > 0.$$

The idea of the proof is to choose $\delta = \delta(M, \sigma^2, \lambda_1, \beta)$ in some specific way in order to guarantee that the deterministic compensation term in (6.7) is lower bounded by $-\lambda_1^2 \delta$, and to pick afterwards some suitable projection in dependence of C and \mathbf{E} out of this class which realizes the lower bound in expectation. Since $\tilde{\pi}_{r,U} = U' \tilde{\pi}_r U \in \mathcal{S}_{M,r}$ for any $\tilde{\pi}_r \in \mathcal{S}_{M,r}$ and $U \in O(M)$, $\check{\mathbf{E}}_{U,V} = U' \mathbf{E} V =_{\mathcal{D}} \mathbf{E}$ for any two fixed orthogonal matrices U and V , and $\|U' A V\|_{S_2}^2 = \|A\|_{S_2}^2$ for any $A \in \mathbb{R}^{M \times M}$, i.e.,

$$\|\tilde{\pi}_r(U \Lambda V' + \mathbf{E})\|_{S_2}^2 = \|\tilde{\pi}_{r,U}(\Lambda + \check{\mathbf{E}}_{U,V})\|_{S_2}^2,$$

we may and do assume subsequently without loss of generality that $U = V = \text{Id}$ and $\pi_1 = \text{Id}_1$ in particular. With e_1, e_2, \dots, e_M denoting the canonical basis vectors of \mathbb{R}^M , every projection matrix in $\mathcal{S}_{M,1}$ can be written as

$$\left(\sum_{i=1}^M \gamma_i e_i \right) \left(\sum_{i=1}^M \gamma_i e_i' \right) =: \tilde{\pi}_{1,\gamma}, \quad \text{with} \quad \sum_{i=1}^M \gamma_i^2 = 1.$$

In order bound $\delta_{C,M,\sigma^2,1}$ from below by means of (6.6) – (6.7), define

$$\delta_{**} := d \frac{(M-1)\sigma^2}{2\lambda_1^2}$$

for some constant $d \in (0, \min(\beta^2, 1))$ to be chosen later. Note that d shall be chosen independently of M, σ^2, λ_1 and C , but is allowed to depend on β only. Furthermore, $\delta^* \leq 1/2$ because $\lambda_1 \geq \beta\sigma\sqrt{M}$. Now, since

$$\tilde{\pi}_{1,\gamma} \text{Id}_1 = \begin{pmatrix} \gamma_1^2 & 0 & \cdots & 0 \\ \gamma_1\gamma_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ \gamma_1\gamma_M & 0 & \cdots & 0 \end{pmatrix},$$

the constraint

$$\tilde{\pi}_{1,\gamma} \in \mathcal{S}_{M,1}(\delta_{**}) \Leftrightarrow \|\tilde{\pi}_{1,\gamma} - \text{Id}_1\|_{S_2}^2 \leq 2\delta_{**} \Leftrightarrow \|(\text{Id} - \tilde{\pi}_{1,\gamma})\text{Id}_1\|_{S_2}^2 \leq \delta_{**}$$

translates into

$$\left((1 - \gamma_1^2)^2 + \sum_{i=2}^M \gamma_1^2 \gamma_i^2 = (1 - \gamma_1^2)^2 + \gamma_1^2(1 - \gamma_1^2) \right) \quad 1 - \gamma_1^2 \leq \delta_{**}.$$

With the choice

$$\gamma_1^* := \sqrt{1 - \delta_{**}} \quad \text{and} \quad \gamma_i^* := \text{sign}(\text{E}_{i1}) \delta_{**}^{1/2} / \sqrt{M-1} \quad \text{for } i = 2, \dots, M,$$

it holds that $\|\tilde{\pi}_{1,\gamma^*} - \text{Id}_1\|_{S_2}^2 \leq 2\delta_{**}$, i.e. $\tilde{\pi}_{1,\gamma^*}$ belongs to $\mathcal{S}_{M,1}(\delta_{**})$. Together with $\delta = \delta_{**}$ in (6.6) – (6.7), this yields

$$(6.8) \quad \delta_{C,M,\sigma^2,1} \geq \mathbf{E} \left(\|\tilde{\pi}_{1,\gamma^*} \text{E}\|_{S_2}^2 - \|\text{Id}_1 \text{E}\|_{S_2}^2 + 2 \text{tr} \left(\text{E}'(\tilde{\pi}_{1,\gamma^*} - \text{Id}_1)C \right) - d\sigma^2(M-r) \right).$$

We evaluate the expressions within the expectation separately. For any $1 \leq i, j \leq M$,

$$(\tilde{\pi}_{1,\gamma} \text{E})_{ij} = \sum_{l=1}^M \gamma_i \gamma_l \text{E}_{lj}.$$

Since $\sum_{i=1}^M \gamma_i^{*2} = 1$, $\text{sign}(\mathbf{E}_{i1})$ and $|\mathbf{E}_{i1}|$ are stochastically independent, and \mathbf{E}_{ij} , $1 \leq i, j \leq M$, are independent by assumption with $\mathbf{E}(\mathbf{E}_{ij}) = 0$ and $\mathbf{E} \text{sign}(\mathbf{E}_{ij}) = 0$,

$$\begin{aligned} \mathbf{E} \|\tilde{\pi}_{1,\gamma^*} \mathbf{E}\|_{S_2}^2 &= \sum_{i,j=1}^M \mathbf{E} \left(\sum_{l=1}^M \gamma_i^* \gamma_l^* \mathbf{E}_{lj} \right)^2 \\ &= \sum_{i,j=1}^M \sum_{l=1}^M \gamma_i^{*2} \gamma_l^{*2} \mathbf{E}(\mathbf{E}_{lj}^2) + \sum_{\substack{l,l' \geq 2: \\ l \neq l'}}^M |\gamma_l^* \gamma_{l'}^*| \mathbf{E} |\mathbf{E}_{l1} \mathbf{E}_{l'1}| \\ &= M\sigma^2 + \sum_{\substack{l,l' \geq 2: \\ l \neq l'}}^M |\gamma_l^* \gamma_{l'}^*| \mathbf{E} |\mathbf{E}_{l1} \mathbf{E}_{l'1}|. \end{aligned}$$

Therefore,

$$(6.9) \quad \mathbf{E} \left(\|\tilde{\pi}_{1,\gamma^*} \mathbf{E}\|_{S_2}^2 - \|\text{Id}_1 \mathbf{E}\|_{S_2}^2 \right) \geq 0.$$

Next, we decompose

$$(6.10) \quad 2 \text{tr}(\mathbf{E}' \tilde{\pi}_{1,\gamma^*} C) = 2 \text{tr}(\mathbf{E}' \tilde{\pi}_{1,\gamma^*} \text{Id}_1 C) + 2 \text{tr}(\mathbf{E}' \tilde{\pi}_{1,\gamma^*} (\text{Id} - \text{Id}_1) C).$$

In order to check $\mathbf{E} \text{tr}(\mathbf{E}' \tilde{\pi}_{1,\gamma^*} (\text{Id} - \text{Id}_1) C) = 0$, it is sufficient to notice that all entries of the matrix $\mathbf{E}' \tilde{\pi}_{1,\gamma^*} (\text{Id} - \text{Id}_1)$ have expectation 0. Indeed, its first column is equal to zero, and one easily verifies for the remaining indices $1 \leq i \leq M$, $2 \leq j \leq M$ that $\mathbf{E}(\mathbf{E}_{li} \gamma_j^* \gamma_l^*) = 0$ for every $1 \leq l \leq M$, hence

$$\left(\mathbf{E}(\mathbf{E}' \tilde{\pi}_{1,\gamma^*} (\text{Id} - \text{Id}_1)) \right)_{ij} = \sum_{l=1}^M \mathbf{E}(\mathbf{E}_{li} \gamma_j^* \gamma_l^*) = 0.$$

Together with (6.9) and $\text{Id}_1 C = \lambda_1 \text{Id}_1$, (6.8) reduces to

$$\begin{aligned} \delta_{C,M,\sigma^2,1} &\geq \mathbf{E} \left(2 \text{tr}(\mathbf{E}' (\tilde{\pi}_{1,\gamma^*} - \text{Id}_1) \text{Id}_1 C) - d\sigma^2(M-1) \right) \\ &= \mathbf{E} \left(2\lambda_1 \text{tr}(\mathbf{E}' \tilde{\pi}_{1,\gamma^*} \text{Id}_1) - d\sigma^2(M-1) \right) \\ &= \sigma \cdot 2\lambda_1 \mathbf{E} \left(\gamma_1^{*2} (\mathbf{E}_{11}/\sigma) + \sum_{i=2}^M \gamma_1^* \gamma_i^* (\mathbf{E}_{i1}/\sigma) \right) - d\sigma^2(M-1) \\ &= (1 - \delta_{**})^{1/2} \delta_{**}^{1/2} (M-1)^{-1/2} \sigma 2\lambda_1 \sum_{i=2}^M \mathbf{E} |\mathbf{E}_{i1}/\sigma| - d\sigma^2(M-1) \\ &\geq \sqrt{d} \sigma^2 (M-1) \frac{2}{\sqrt{2\pi}} - d\sigma^2(M-1). \end{aligned}$$

Choosing now some $d \in (0, \min(\beta^2, 1))$ such that $2\sqrt{d}/\sqrt{2\pi} - d > 0$ proves (6.5). \square

REMARK. Theorem 6.1 answers to question (A) from the Introduction in the negative.

6.2. A partial solution to the conjecture (6.3) for $r > 1$ in the large amplitude regime. The specific construction of the (random) projection $\tilde{\pi}_{1,\gamma^*}$ in the proof from the previous paragraph cannot canonically be extended to arbitrary $r > 1$. For the result in this subsection, we use finally a different approach based on abstract lower bounds on suprema of Gaussian processes which applies to any $r \leq M - r$. For the prototype $C_\alpha = \alpha \text{Id}_r$ of high accuracy in the large amplitude regime $|\alpha| \gg \sigma\sqrt{M}$ of Theorem 5.1, the following result is deduced. Its substantially more involved extension to a non-asymptotic optimal lower bound for general matrices C may also involve the Sudakov-type minoration for Gaussian chaos processes (Talagrand (1992)).

THEOREM 6.2. *Let $(E_{ij})_{i,j=1}^M$ be a centered matrix of independent Gaussian entries with variance σ^2 . Let $C_{\alpha,s} \in \mathbb{R}^{M \times M}$ with singular value decomposition $U\Lambda_{\alpha,s}V'$, where $\Lambda_{\alpha,s} = \alpha \text{Id}_s$ with $0 < \alpha \in \mathbb{R}$ and $1 \leq s < M$. Then*

$$(6.11) \quad \liminf_{\alpha \rightarrow \infty} \max_{s \in \{r, M-r\}} \delta_{C_{\alpha,s}, M, \sigma^2, s} \gtrsim \sigma^2 r(M-r).$$

The proof of Theorem 6.2 is deferred to Section 9.

REMARK. In view of the polar decomposition of $(\text{Id} - \tilde{\pi}_r)\pi_r$ and $\tilde{\pi}_r(\text{Id} - \pi_r)$ which shows in particular that these two matrices have the same singular values, we conjecture that the bound holds true also without the maximum over $s \in \{r, M-r\}$, but do not have a rigorous and elegant proof for it yet. Note that this maximum is redundant if $r = M/2$, $M \in 2\mathbb{N}$.

7. Consequences on statistical estimation problems. Let $X = C + E$ as described in (1.3). Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_M$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ denote the singular values of X and C , respectively. Recall that $\sum_{i=1}^r \hat{\lambda}_i^2 = \|\hat{\pi}_r X\|_{S_2}^2$ and $\sum_{i=1}^r \lambda_i^2 = \|\pi_r C\|_{S_2}^2$ with the rank- r -projections $\hat{\pi}_r$ and π_r as defined in the Introduction.

7.1. The largest singular value. We begin with the simplest example of estimating λ_1^2 , the largest eigenvalue of $C'C$, based on the observation $X = C + E$. As explained in the introduction, the maximal eigenvalue of $X'X$ is positively biased as an estimator for λ_1^2 , because

$$\mathbf{E} \hat{\lambda}_1^2 = \mathbf{E} \|\hat{\pi}_1 X\|_{S_2}^2 \geq \mathbf{E} \|\pi_1 X\|_{S_2}^2 = \lambda_1^2 + \sigma^2 M.$$

Therefore, one immediate improvement is to consider $\hat{s} := \hat{\lambda}_1^2 - \sigma^2 M$ as an estimator for λ_1^2 . As Theorem 6.1 reveals for the particular case $r = 1$,

$$(7.1) \quad \mathbf{E} \hat{s} - \lambda_1^2 = \mathbf{E} \hat{\lambda}_1^2 - \sigma^2 M - \lambda_1^2$$

is strictly positive and bounded away from zero, uniformly over $C \in \mathbb{R}^{M \times M}$. As a consequence of Corollary 4.2 and Theorem 6.1,

$$\delta_{C, M, \sigma^2, 1} = \mathbf{E} \hat{s} - \lambda_1^2 \in \left[c_1 \sigma^2 M, c_2 (\sigma^2 M + \sigma \sqrt{M} \|C\|_{S_\infty}) \right]$$

for some universal real constants $c_1, c_2 > 0$ which do not depend on M, σ^2 and C , and it follows from (1.2) for $C = 0$ and Proposition 3.1 (ii) for $C = \alpha \text{Id}$ that these bounds cannot be improved in general. Hence, one message of our analysis is:

The quantity $\sigma^2 M$ always underestimates the bias $\mathbf{E}(\hat{\lambda}_1^2 - \lambda_1^2)$ by at least some universal factor strictly larger than 1, independently on how favorable the matrix C is.

In other words, even after correction by $\sigma^2 M$, the difference

$$(7.2) \quad \mathbf{E} \hat{\lambda}_1^2 - \lambda_1^2 - \sigma^2 M$$

remains of the same order $\sigma^2 M$ at least, independently of C . Moreover, there exist matrices C for which (7.1) is not smaller in order than $\sigma^2 M + \sigma \sqrt{M} \|C\|_{S_\infty}$. That is, large amplitude $\|C\|_{S_\infty}$ never improves (in order) the accuracy as compared to $C = 0$, but it may result in substantially worse accuracy of approximation. Therefore, some further consequence is that small magnitude of $\sigma^2 M$ is necessary but far from being sufficient for the bias of \hat{s} in (7.1) to be small. The worst case error is non-asymptotically quantified in terms of $\|C\|_{S_\infty}, \sigma^2$ and M in Corollary 4.2 (i). Theorem 5.1 describes more precisely the effect of the shape of the singular value spectrum on the accuracy of approximation. For example, if $C = \alpha \text{Id}$, then (7.2) grows like $|\alpha| \sigma \sqrt{M}$ as $|\alpha| \rightarrow \infty$, cf. Proposition 3.1 (ii), for any fixed σ^2 and M . On the other hand, if $C = \alpha \text{Id}_1$, then (7.2) remains bounded by some universal constant times $\sigma^2 M$, independently of α . The same holds true for the full rank matrix $C = \alpha \text{Id}_1 + \alpha \text{Id}$. Consequently, Theorem 5.1 demonstrates that large amplitude $\|C\|_{S_\infty}$ does not necessarily result in worse accuracy of approximation as compared to the case $C = 0$, and discovers some prototypes of high accuracy in the large amplitude regime. Similar conclusions for $r > 1$, i.e. statistics of the form $\sum_{i=1}^r \hat{\lambda}_i^2$, are valid as well.

7.2. Relative quantities. This subsection is devoted to the consequences of our results on relative quantities as described in the introduction. Consider, for instance, the ratio

$$t_r := \frac{\|\pi_r C\|_{S_2}^2}{\|C\|_{S_2}^2} = \frac{\sum_{i=1}^r \lambda_i^2}{\sum_{i=1}^M \lambda_i^2}.$$

Assuming $\|C\|_{S_2}^2$ to be known, some natural candidate estimator of t_r is

$$\hat{t}_r := \frac{\|\hat{\pi}_r X\|_{S_2}^2 - \sigma^2 r M}{\|C\|_{S_2}^2}.$$

In this situation, increasing amplitude of C always results in smaller bias of the estimator \hat{t}_r . Suppose that $C = \alpha A$ for some real number $\alpha \in \mathbb{R}$ and an $M \times M$ -matrix A with $\|A\|_{S_\infty} = 1$. Note that $\|A\|_{S_2}^2 \geq \|A\|_{S_\infty}^2 = 1$. Then Corollary 4.2 (i) yields

$$\sup_{\substack{A \in \mathbb{R}^{M \times M}: \\ \text{rank}(A) \geq r, \|A\|_{S_\infty} = 1}} \mathbf{E}(\hat{t}_r - t_r) \lesssim \frac{\sigma^2 r M}{\alpha^2} \left\{ 1 + \frac{|\alpha|}{\sigma \sqrt{M}} \right\}.$$

For every fixed A , σ^2 , r and M , the bias $\mathbf{E}(\widehat{t}_r - t_r)$ tends to zero as $|\alpha| \rightarrow \infty$, in contrast to the situation for the absolute difference in the previous Subsection 7.1. Note that it decreases of the order $|\alpha|^{-1}$ at least and of the order α^{-2} at most, depending on the shape of the singular value spectrum of A , cf. Theorem 5.1. So, whereas, for every fixed A , σ^2 , r and M , the *absolute difference* as described in Subsection 7.1 for $r = 1$ cannot get closer to zero as the amplitude $|\alpha|$ of $C = \alpha A$ increases, independently on how favorable the matrix C may be, the *relative difference* $\mathbf{E}(\widehat{t}_r - t_r)$ tends to zero as the amplitude goes to infinity for every matrix $C = \alpha A$. The shape of the singular value spectrum however clearly influences the accuracy of approximation, in the same fashion as explained in Subsection 7.1, cf. Theorem 5.1.

7.3. Quadratic functionals of low-rank matrices. One natural candidate for estimating $\|C\|_{S_2}^2$, based on the observation $X = C + E$ described by (1.3), is the unbiased estimator $\|X\|_{S_2}^2 - \sigma^2 M^2$. Simple calculation yields

$$(7.3) \quad \text{Var}\left(\|X\|_{S_2}^2 - \sigma^2 M^2\right) = 2\sigma^4 M^2 + 4\sigma^2 \|C\|_{S_2}^2.$$

One disadvantage of this estimator is its large variance for large values of M : it depends quadratically on the dimension. If $r = \text{rank}(C) < M$, then the matrix C can be fully characterized by $(2M - r)r$ parameters as can be seen by the singular value decomposition. That is, if $r \ll M$, the intrinsic dimension of the problem is of the order rM rather than M^2 . Now observe that for every matrix C with $\text{rank}(C) = r$,

$$\|C\|_{S_2}^2 = \|\pi_r C\|_{S_2}^2.$$

Elementary calculation reveals that $\|\pi_r X\|_{S_2}^2 - \sigma^2 rM$ unbiasedly estimates $\|C\|_{S_2}^2$, and

$$(7.4) \quad \text{Var}\left(\|\pi_r X\|_{S_2}^2 - \sigma^2 rM\right) = 2\sigma^4 rM + 4\sigma^2 \|C\|_{S_2}^2.$$

As compared to (7.3), its variance does not depend on the squared dimension M^2 but grows like rM , which can be substantially smaller. Moreover,

$$\mathbf{E}\left(\|\pi_r X\|_{S_2}^2 - \sigma^2 rM - \|C\|_{S_2}^2 - 2\sigma \text{tr}(E'C)\right)^2 = 2\sigma^4 rM,$$

that is, $\sigma^{-1}(\|\pi_r X\|_{S_2}^2 - \sigma^2 rM - \|C\|_{S_2}^2)$ is approximately centered normal with variance $4\|C\|_{S_2}^2$ if $\sigma^2 rM = o(1)$ in an asymptotic framework, and $4\|C\|_{S_2}^2$ is the asymptotic efficiency lower bound (Laurent and Massart (2000)). The statistics $\|\pi_r X\|_{S_2}^2 - \sigma^2 rM$, however, cannot be used for estimating $\|C\|_{S_2}^2$ since $\pi_r = \pi_r(C)$ depends on C itself and is unknown a priori. Unfortunately, Theorem 6.1 and Theorem 6.2 demonstrate that the same result cannot be shown with $\|\widehat{\pi}_r X\|_{S_2}^2 - \sigma^2 rM$ in place of $\|\pi_r X\|_{S_2}^2 - \sigma^2 rM$, because the bias $\mathbf{E}\|\widehat{\pi}_r X\|_{S_2}^2 - \mathbf{E}\|\pi_r X\|_{S_2}^2$ is of the order not smaller than $\sigma^2 r(M - r)$, i.e.

$$\sigma^{-1}(\mathbf{E}\|\widehat{\pi}_r X\|_{S_2}^2 - \mathbf{E}\|\pi_r X\|_{S_2}^2)$$

is not negligible under the same conditions, even not for very favorable matrices $C = \alpha \text{Id}_r$. That is, empirical low-rank projections $\|\widehat{\pi}_r X\|_{S_2}^2 - \sigma^2 r M$ cannot be successfully used for efficient estimation of $\|C\|_{S_2}^2$, even if the $\text{rank}(C) \ll M$ is explicitly known beforehand. Note that in contrast, empirical low-rank approximations have been proved useful when estimating a low-rank matrix C under Hilbert-Schmidt norm loss, see [Bunea et al. \(2011\)](#), [Koltchinskii \(2011\)](#), [Negahban and Wainwright \(2011\)](#), and [Rohde and Tsybakov \(2011\)](#).

The problem of quadratic functional estimation in the matrix context appears, for instance, in the recent area of low-rank matrix recovery, when one is interested in recovering the linear entropy $1 - \|C\|_{S_2}^2$ of a quantum density matrix C as an approximation of von Neumann entropy based on noisy observations. We refer the reader to [Artiles et al. \(2004\)](#) for a detailed description of applications in quantum state tomography and the recent article of [Koltchinskii \(2011\)](#) for low-rank matrix recovery of quantum density matrices. Note however that our results do not take into account that a quantum density matrix C is self-adjoint, and the Wigner ensemble may behave differently as already outlined in the introduction. In view of model selection issues, an estimate of the bias is even required over the whole scale $r \in \{1, \dots, M\}$ since the rank is typically unknown a priori and low at most approximately, for which reason exact asymptotic results for uniformly bounded rank perturbations are of limited value for this application.

8. Proof of Theorem 5.1. This section is devoted to the proof of our main result. Subsection 8.1 deals with the description of the sets $\mathcal{G}_{M,r}(\delta) := \{\widetilde{\pi}_r \in \mathcal{S}_{M,r} : \|\pi_r C\|_{S_2}^2 - \|\widetilde{\pi}_r C\|_{S_2}^2 \leq \delta\}$, which characterize the slices $\mathcal{A}_{C,k} = \mathcal{G}_{M,r}(2^{-k} \|\pi_r C\|_{S_2}^2) \setminus \mathcal{G}_{M,r}(2^{-(k+1)} \|\pi_r C\|_{S_2}^2)$. It is shown that these sets can be approximated by sets of simple geometric structure. In case of some substantial spectral gap in r , this approximation is very tight. Sharp bounds on their metric entropy are derived in Subsection 8.2. The final arguments are given in Subsection 8.3.

8.1. Characterizing the sets $\mathcal{G}_{M,r}(\delta)$. The first goal for a sophisticated analysis is to characterize the sets

$$\mathcal{G}_{M,r}(\delta, C) := \left\{ \widetilde{\pi}_r \in \mathcal{S}_{M,r} : \|\pi_r C\|_{S_2}^2 - \|\widetilde{\pi}_r C\|_{S_2}^2 \leq \delta \right\}, \quad \delta > 0.$$

Note that

$$\mathcal{A}_{C,k} = \mathcal{G}_{M,r}(2^{-k} \|\pi_r C\|_{S_2}^2) \setminus \mathcal{G}_{M,r}(2^{-(k+1)} \|\pi_r C\|_{S_2}^2).$$

This is a very delicate part and quite involved in general, but we find below a tight characterization in case of approximate rank- r -matrices, i.e. those matrices for which $\|(\text{Id} - \pi_r)C\|_{S_2}$ is small. For reasons of clarity, we restrict attention to matrices of rank larger or equal to r .

PROPOSITION 8.1. *Let $C \in \mathbb{R}^{M \times M}$, $\text{rank}(C) \geq r$, with singular values $\lambda_i, i = 1, \dots, M$, ordered in decreasing magnitude. Denote $\Delta_r^* := \sum_{i=r+1}^{2r} \lambda_i^2$*

and $\gamma_r^* := (\lambda_r^2 - \lambda_{r+1}^2)^{-1/2}$ if $\lambda_r > \lambda_{r+1}$, and $\gamma^* = \infty$ else. Then for any $\pi_r \in \text{Arg max}_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \|\tilde{\pi}_r C\|_{S_2}^2$,

$$(8.1) \quad \mathcal{G}_{M,r}(\delta, C) \subset \left\{ \tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\pi_r - \tilde{\pi}_r\|_{S_2} \leq \min \left(\lambda_r^{-1} \sqrt{2(\delta + \Delta_r^*)}, \gamma_r^* \sqrt{2\delta} \right) \right\}$$

and

$$(8.2) \quad \left\{ \tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\pi_r - \tilde{\pi}_r\|_{S_2} \leq \lambda_1^{-1} \sqrt{2\delta} \right\} \subset \mathcal{G}_{M,r}(\delta, C).$$

PROOF. Let $U\Lambda V'$ denote some singular value decomposition of C , with

$$\Lambda_r := \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r, 0, \dots, 0) \quad \text{and} \quad \Lambda_{M-r} := \Lambda - \Lambda_r.$$

Recall that $\pi_r := \sum_{i=1}^r U_i U_i'$ is some maximizer of $\|\tilde{\pi}_r C\|_{S_2}^2$ over $\tilde{\pi}_r \in \mathcal{S}_{M,r}$, where the U_i 's denote the column vectors of U . Note at this point that due to multiplicities in the singular value spectrum, the orthonormal vectors U_1, \dots, U_r are not unique in general. As concerns the proof of (8.1), we check first that

$$(8.3) \quad \|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \geq \lambda_r^2 \frac{1}{2} \|\pi_r - \tilde{\pi}_r\|_{S_2}^2 - \Xi_r$$

for all $\tilde{\pi}_r \in \mathcal{S}_{M,r}$, where $\Xi_r := \|\tilde{\pi}_r U \Lambda_{M-r} U'\|_{S_2}^2 = \|(\text{Id} - \pi_r) \tilde{\pi}_r (\text{Id} - \pi_r) U \Lambda_{M-r} U'\|_{S_2}^2$. Using the symmetry of the projection matrices π_r and $\tilde{\pi}_r$ and the invariance of the trace operator under cyclic permutation, we obtain the identity

$$(8.4) \quad \begin{aligned} \|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 &= \text{tr} \left(C' (\pi_r' \pi_r - \tilde{\pi}_r' \tilde{\pi}_r) C \right) \\ &= \text{tr} \left(C C' (\pi_r - \tilde{\pi}_r) \right) \\ &= \text{tr} \left(\Lambda_r^2 U' (\pi_r - \pi_r \tilde{\pi}_r \pi_r) U \right) - \text{tr} \left(\Lambda_{M-r}^2 U' (\text{Id} - \pi_r) \tilde{\pi}_r (\text{Id} - \pi_r) U \right) \end{aligned}$$

$$(8.5) \quad = \sum_{i=1}^r \lambda_i^2 [U' (\pi_r (\text{Id} - \tilde{\pi}_r) \pi_r) U]_{ii}$$

$$(8.6) \quad - \sum_{i=r+1}^M \lambda_i^2 [U' (\text{Id} - \pi_r) \tilde{\pi}_r (\text{Id} - \pi_r) U]_{ii}$$

Note that the sum in (8.6) equals Ξ_r . Since $\pi_r - \pi_r \tilde{\pi}_r \pi_r = \pi_r (\text{Id} - \tilde{\pi}_r) \pi_r = \pi_r' (\text{Id} - \tilde{\pi}_r) \pi_r$ is positive semidefinite, all summands in the first term of (8.5) are non-

negative. Consequently,

$$\begin{aligned}
 (8.5) &\geq \lambda_r^2 \sum_{i=1}^r [U'(\pi_r - \pi_r \tilde{\pi}_r \pi_r)U]_{ii} \\
 (8.7) &= \lambda_r^2 \sum_{i=1}^M [U'(\pi_r - \pi_r \tilde{\pi}_r \pi_r)U]_{ii} \\
 &= \lambda_r^2 \text{tr}(\pi_r - \pi_r \tilde{\pi}_r \pi_r),
 \end{aligned}$$

where (8.7) follows from $[U'(\pi_r - \pi_r \tilde{\pi}_r \pi_r)U]_{ii} = 0$ for $i > r$. By positive semidefiniteness again and symmetry, the eigenvalue decomposition of $\pi_r - \pi_r \tilde{\pi}_r \pi_r$ and the invariance of the trace operator under basis transformation yield $\text{tr}(\pi_r - \pi_r \tilde{\pi}_r \pi_r) = \|\pi_r - \pi_r \tilde{\pi}_r \pi_r\|_{S_1}$, hence

$$\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \geq \lambda_r^2 \|\pi_r - \pi_r \tilde{\pi}_r \pi_r\|_{S_1} - \Xi_r.$$

Now, since $\pi_r = \pi_r \tilde{\pi}_r + \pi_r(\text{Id} - \tilde{\pi}_r)$ is an orthogonal decomposition,

$$\|\pi_r \tilde{\pi}_r\|_{S_2}^2 + \|\pi_r(\text{Id} - \tilde{\pi}_r)\|_{S_2}^2 = \|\pi_r\|_{S_2}^2 = r = \|\tilde{\pi}_r\|_{S_2}^2 = \|\pi_r \tilde{\pi}_r\|_{S_2}^2 + \|(\text{Id} - \pi_r) \tilde{\pi}_r\|_{S_2}^2,$$

implying that $\|\pi_r(\text{Id} - \tilde{\pi}_r)\|_{S_2}^2 = \|(\text{Id} - \pi_r) \tilde{\pi}_r\|_{S_2}^2$. Consequently,

$$\begin{aligned}
 (8.8) \quad \|\pi_r - \pi_r \tilde{\pi}_r \pi_r\|_{S_1} &= \text{tr}(\pi_r - \pi_r \tilde{\pi}_r \pi_r) = \text{tr}(\pi_r' \pi_r - \pi_r' \tilde{\pi}_r \pi_r) \\
 &= \|\pi_r\|_{S_2}^2 - \|\pi_r \tilde{\pi}_r\|_{S_2}^2 \\
 &= \|\pi_r(\text{Id} - \tilde{\pi}_r)\|_{S_2}^2 = \frac{1}{2} \|\pi_r - \tilde{\pi}_r\|_{S_2}^2,
 \end{aligned}$$

where the last equality follows by the orthogonality of the decomposition $\tilde{\pi}_r - \pi_r = (\text{Id} - \pi_r) \tilde{\pi}_r - \pi_r(\text{Id} - \tilde{\pi}_r)$. This implies (8.3). In order to deduce the bound in (8.1), note first that because of the positive semidefiniteness of $(\text{Id} - \pi_r) \tilde{\pi}_r (\text{Id} - \pi_r)$, also all summands in the second sum (8.6) are non-negative, whence

$$\sum_{i=r+1}^M \lambda_i^2 [U(\text{Id} - \pi_r) \tilde{\pi}_r (\text{Id} - \pi_r) U']_{ii} \leq \lambda_{r+1}^2 \sum_{i=r+1}^M [U(\text{Id} - \pi_r) \tilde{\pi}_r (\text{Id} - \pi_r) U']_{ii}.$$

With the same arguments as provided above for $\pi_r - \pi_r \tilde{\pi}_r \pi_r = \pi_r(\text{Id} - \tilde{\pi}_r) \pi_r$, we deduce

$$(8.9) \quad \Xi_r \leq \lambda_{r+1}^2 \frac{1}{2} \|(\text{Id} - \pi_r) - (\text{Id} - \tilde{\pi}_r)\|_{S_2}^2 = \lambda_{r+1}^2 \frac{1}{2} \|\pi_r - \tilde{\pi}_r\|_{S_2}^2.$$

Moreover, because $U' \tilde{\pi}_r U \in \mathcal{S}_{M,r}$ for every $U \in O(M)$,

$$(8.10) \quad \Xi_r = \|\tilde{\pi}_r U \Lambda_{M-r} U'\|_{S_2}^2 = \|U' \tilde{\pi}_r U \Lambda_{M-r}\|_{S_2}^2 \leq \sum_{i=r+1}^{2r} \lambda_i^2,$$

and claim (8.1) follows from (8.3), together with (8.9) and (8.10). The proof of (8.2) uses that the expression (8.5) is in turn upper bounded by $\lambda_1^2 \|\pi_r - \pi_r \tilde{\pi}_r \pi_r\|_{S_1}$ while

(8.6) is less or equal to zero, and concludes finally with the same equality chain (8.8). \square

We note that in the particular case of rank- r -matrices, $\sum_{i=r+1}^{2r} \lambda_i = 0$, i.e. the first term in the upper bound on the radius of the S_2 -ball in (8.1) coincides up to the ratio λ_1/λ_r with the lower bound in (8.2). Equality holds for rank- r -matrices with rectangular singular value spectrum $\lambda_1 = \lambda_2 = \dots = \lambda_r$. If $C = \alpha \text{Id}$ for some $\alpha \neq 0$, then $\Delta_r^* = \alpha^2 r$ and the inclusion (8.1) is trivial:

$$\left\{ \tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\pi_r - \tilde{\pi}_r\|_{S_2} \leq \alpha^{-1} \sqrt{2(\delta + \alpha^2 r)} \right\} = \mathcal{S}_{M,r}, \quad \text{for any } \delta > 0.$$

This is in accordance with the fact that in this case, also $\mathcal{G}_{M,r}(\delta) = \mathcal{S}_{M,r}$ for any $\delta > 0$.

8.2. Metric entropy bounds. The nice feature of the results from the previous paragraph is that they enable us to determine tight bounds on the metric entropy on these particular subsets of the Grassmannian by the volumetric argument. We provide a slightly refined version. Recall at this point the definition of the covering numbers. For any totally-bounded, pseudometric space (\mathcal{X}, d) and any subset $E \subset \mathcal{X}$, the covering number $N(E, d, \delta)$ is the smallest number of closed d -balls in \mathcal{X} of radius δ needed to cover E .

LEMMA 8.2. *For any $\pi_r \in \mathcal{S}_{M,r}$ and $\delta > 0$, let $B_{S_q}(\pi_r, \delta)$ denote the closed S_q -ball with center π_r of radius δ . Then there exist universal constants $\hat{c}, C, c > 0$ such that for all $0 < \Delta \leq r$, $r \leq M - r$ and $0 < \delta \leq \Delta$,*

$$(8.11) \quad \log N\left(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta\right) \leq \min\left(\hat{c}rM \frac{\Delta^2}{\delta^2}, r(M-r) \log\left(\frac{C\Delta}{\delta}\right)\right),$$

$$(8.12) \quad \log N\left(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_\infty}, \delta\right) \leq \min\left(\hat{c}rM \frac{\Delta}{\sqrt{r}\delta}, r(M-r) \log\left(\frac{C\Delta}{\delta}\right)\right)$$

as well as

$$(8.13) \quad \log N\left(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta\right) \geq r(M-r) \log\left(\frac{c\Delta}{\delta}\right) \quad \text{and}$$

$$(8.14) \quad \log N\left(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_\infty}, \delta\right) \geq r(M-r) \log\left(\frac{c\Delta}{\sqrt{r}\delta}\right)$$

REMARK. (i) $\hat{c}^{1/2}$ equals c_{Sud} times a uniform (in M) bound on the expected spectral norm of $E/(\sqrt{M}\sigma)$, where c_{Sud} is the universal constant of Sudakov's minoration which is bounded by 6 (see [Ledoux \(1996\)](#)). C is proportional to the ratio of the constants C'/c' which appear in the bounds of the metric entropy of the Grassmann manifold, see the proof below. Estimates on their values are not provided in [Szarek \(1982\)](#).

(ii) For arbitrary $1 \leq q \leq \infty$, the bound
(8.15)

$$r(M-r) \log \left(\frac{c\Delta}{\delta} \right) \leq \log N \left(B_{S_q}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_q}, \delta \right) \leq r(M-r) \log \left(\frac{C\Delta}{\delta} \right)$$

can be proved completely analogously, replacing below S_2 by S_q .

PROOF. By the geometric formulation of Sudakov's minoration, the trace duality and the Cauchy-Schwarz inequality,

$$\begin{aligned} \delta \sqrt{\log N \left(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta \right)} &\lesssim \mathbf{E} \sup_{T \in B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}} \text{tr}(T(\mathbf{E}/\sigma)) \\ &\leq \mathbf{E} \sup_{T \in B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}} \text{tr}((T - \pi_r)(\mathbf{E}/\sigma)) + \mathbf{E} \text{tr}(\pi_r(\mathbf{E}/\sigma)) \\ &\leq \sup_{T \in B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}} \|T - \pi_r\|_{S_1} \mathbf{E} \|\mathbf{E}/\sigma\|_{S_\infty} \\ &\lesssim \Delta \sqrt{rM}, \end{aligned}$$

which provides the first estimate in the minimum of (8.11). In order to prove the second term, note first that $N(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta)$ does not depend on the specific $\pi_r \in \mathcal{S}_{M,r}$. Similarly to the covering number, the capacity number $D(E, d, \delta)$ is the largest number of elements of E having distance d strictly larger than δ to each other. Using the relations between covering and packing (capacity) numbers (cf. Theorem 1.2.1, [Dudley \(1999\)](#)),

$$\begin{aligned} N(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta) N(\mathcal{S}_{M,r}, d_{S_2}, 4\Delta) \\ \leq D(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta) D(\mathcal{S}_{M,r}, d_{S_2}, 4\Delta). \end{aligned}$$

Let $\{\pi_r^{(1)}, \dots, \pi_r^{(k_{4\Delta})}\}$ be some maximal subset of $\mathcal{S}_{M,r}$ with $d_{S_2}(\pi_r^{(l)}, \pi_r^{(m)}) > 4\Delta$ for all $l \neq m \in \{1, \dots, k_{4\Delta}\}$, $k_{4\Delta} = D(\mathcal{S}_{M,r}, d_{S_2}, 4\Delta)$. Then

$$\sum_{j=1}^{k_{4\Delta}} D(B_{S_2}(\pi_r^{(j)}, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta) \leq D(\mathcal{S}_{M,r}, d_{S_2}, \delta/2) \leq N(\mathcal{S}_{M,r}, d_{S_2}, \delta/4),$$

that is $N(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta) \leq N(\mathcal{S}_{M,r}, d_{S_2}, \delta/4) / N(\mathcal{S}_{M,r}, d_{S_2}, 4\Delta)$.

Now (8.11) follows by an immediate application of Proposition 8, [Pajor \(1998\)](#), which states

$$\left(\frac{c'}{\xi} \right)^{r(M-r)} \leq N(\mathcal{S}_{M,r}, d_{S_q}, \xi r^{1/q}) \leq \left(\frac{C'}{\xi} \right)^{r(M-r)}$$

for $1 \leq q \leq \infty$ and universal constants $c', C' > 0$. As concerns bound (8.12), first observe that

(8.16)

$$N(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_\infty}, \delta) \leq N(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta/\theta) N(B_{S_2}(0, 1), d_{S_\infty}, \theta)$$

for any $\theta > 0$. Combining (8.16) with (8.11) and the bound $\log N(B_{S_2}(0, 1), d_{S_\infty}, \delta) \lesssim M\delta^{-2}$ (cf. Pajor 1998, Lemma 4), we obtain

$$\log N(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_\infty}, \delta) \lesssim rM \frac{\Delta^2 \theta^2}{\delta^2} + M\theta^{-2}.$$

Choosing $\theta^2 = \delta \Delta^{-1} r^{-1/2}$ gives the first term in the minimum on the RHS of (8.12). The proof of the second bound in (8.12) follows from (8.11) since $d_{S_\infty} \leq d_{S_2}$. As concerns the reverse inequality (8.13), it is sufficient to note that

$$D(\mathcal{S}_{M,r}, d_{S_2}, \Delta) N(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta) \geq N(\mathcal{S}_{M,r}, d_{S_2}, \delta)$$

which after applying the inequalities of Theorem 1.2.1 in Dudley (1999) again is lower bounded by

$$N(B_{S_2}(\pi_r, \Delta) \cap \mathcal{S}_{M,r}, d_{S_2}, \delta) \geq \frac{N(\mathcal{S}_{M,r}, d_{S_2}, \delta)}{N(\mathcal{S}_{M,r}, d_{S_2}, \Delta/2)},$$

and the result follows as above by an application of Proposition 8, Pajor (1998). (8.14) follows analogously from $d_{S_\infty} \geq (2r)^{-1/2} d_{S_2}$ on $\mathcal{S}_{M,r}$. \square

8.3. Slicing the Grassmann manifold. As has been seen in Section 6, the bound involves some term of the order $\sigma^2 r(M - r)$ at least. Since Proposition 4.1 yields in case $C = 0$ the bound $\delta_{0,M,\sigma^2,r} \lesssim \sigma^2 rM$, we decompose the supremum

$$\begin{aligned} & \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left(\|\hat{\pi}_r(C + E)\|_{S_2}^2 - \|\pi_r(C + E)\|_{S_2}^2 \right) \\ &= \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\{ \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2 + 2 \operatorname{tr}(E'(\tilde{\pi}_r - \pi_r)C) \right. \\ & \quad \left. - \left(\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \right) \right\} \\ &\leq \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\{ \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2 \right\} \\ & \quad + \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\{ 2 \operatorname{tr}(E'(\tilde{\pi}_r - \pi_r)C) - \left(\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \right) \right\} \end{aligned}$$

and treat these two suprema separately. Since

$$\mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\{ \|\tilde{\pi}_r E\|_{S_2}^2 - \|\pi_r E\|_{S_2}^2 \right\} \right) \lesssim \sigma^2 rM$$

by Proposition 4.1 applied to the situation $C = 0$, it is sufficient to prove that the expectation of

$$W := \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left\{ 2 \operatorname{tr}(E'(\tilde{\pi}_r - \pi_r)C) - \left(\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \right) \right\}$$

satisfies the bound of Theorem 5.1. For any $C \in \mathbb{R}^{M \times M}$, $\text{rank}(C) \geq r$, let $C = U\Lambda V'$ denote some singular value decomposition, with diagonal matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$. The singular values are assumed to be ordered in decreasing magnitude. Let

$$\mathcal{S}_{M,r}(\delta) := \left\{ \tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\tilde{\pi}_r - \pi_r\|_{S_2} \leq \sqrt{2\delta} \right\}.$$

In view of the inclusion provided by Proposition 8.1, we deviate slightly from the description at the beginning of Section 5 and conduct the proof of Theorem 5.1 along two different decompositions of $\mathcal{S}_{M,r}$. We shall decompose $\mathcal{S}_{M,r}$ into slices

$$\mathcal{C}_{C,i}(\tilde{\Delta}_k) = \mathcal{B}_{C,i}(\tilde{\Delta}_k) \setminus \mathcal{B}_{C,i}(\tilde{\Delta}_{k-1})$$

along a geometric grid $\tilde{\Delta}_k = 2^{-k+2}r$ up to $k < k_{0,i}$ with $k_{0,i}$ specified below. First, we take

$$(8.17) \quad \mathcal{B}_{C,1}(\tilde{\Delta}_k) := \left\{ \tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \geq \tilde{\Delta}_k \lambda_r^2 - \sum_{i=r+1}^{2r} \lambda_i^2 \right\}.$$

In a second step, whenever $\lambda_r > \lambda_{r+1}$, we choose

$$(8.18) \quad \mathcal{B}_{C,2}(\tilde{\Delta}_k) := \left\{ \tilde{\pi}_r \in \mathcal{S}_{M,r} : \|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \geq (\lambda_r^2 - \lambda_{r+1}^2) \tilde{\Delta}_k \right\}.$$

By Proposition 8.1, $\mathcal{C}_{C,i}(\tilde{\Delta}_k) \subset \mathcal{S}_{M,r}(2\tilde{\Delta}_k)$. Recall that by construction, $\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \geq \tilde{\Delta}_k \lambda_r^2 - \sum_{i=r+1}^{2r} \lambda_i^2$ if $\tilde{\pi}_r \in \mathcal{C}_{C,1}(\tilde{\Delta}_k)$. Whenever $\lambda_{r+1} > 0$, define

$$(8.19) \quad k_{0,1} := \arg \max_{k \in \mathbb{N}_0} \left\{ \tilde{\Delta}_k \lambda_r^2 - \sum_{i=r+1}^{2r} \lambda_i^2 \geq \frac{1}{2} \tilde{\Delta}_k \lambda_r^2 \right\}, \text{ and set } k_{0,1} := \infty \text{ if } \lambda_{r+1} = 0.$$

Define $k_{0,2} := \infty$. Denoting

$$\begin{aligned} W_{k,i} &:= \sup_{\tilde{\pi}_r \in \mathcal{C}_{C,i}(\tilde{\Delta}_k)} \left\{ 2 \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) C \right) - \left(\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \right) \right\}, \\ W_{k,i}^0 &:= \sup_{\tilde{\pi}_r \in \mathcal{C}_{C,i}(\tilde{\Delta}_k)} 2 \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) C \right) \end{aligned}$$

and

$$\widetilde{W}_{k_{0,1}} := \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}(\tilde{\Delta}_{k_{0,1}})} 2 \text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) C \right)$$

with $\tilde{\Delta}_\infty := 0$, we obtain the expansions

$$(8.20) \quad \mathbf{E} W \leq \begin{cases} \sum_{k < k_{0,1}} \mathbf{E} \left(0 \vee (W_{k,1}^0 - \Omega_{k,1}) \right) + \mathbf{E} \widetilde{W}_{k_{0,1}} \\ \sum_{k \in \mathbb{N}} \mathbf{E} \left(0 \vee (W_{k,2}^0 - \Omega_{k,2}) \right) \end{cases}$$

where $\Omega_{k,1} = \tilde{\Delta}_k \lambda_r^2 - \sum_{i=r+1}^{2r} \lambda_i^2$ and $\Omega_{k,2} = (\lambda_r^2 - \lambda_{r+1}^2) \tilde{\Delta}_k$. Note that $\mathbf{E} \widetilde{W}_\infty = 0$, that is, $\mathbf{E} \widetilde{W}_{k_{0,1}} = 0$ if $\lambda_{r+1} = 0$.

PROOF OF THEOREM 5.1. Note that each $W_{k,i}^0$ is the supremum over some Gaussian process, and

$$\sup_{\tilde{\pi}_r \in \mathcal{C}_{C,i}(\tilde{\Delta}_k)} \text{Var} \left(\text{tr} \left(\mathbf{E}'(\tilde{\pi}_r - \pi_r) C \right) \right) \leq \sup_{\tilde{\pi}_r \in \mathcal{C}_{C,i}(\tilde{\Delta}_k)} \sigma^2 \lambda_1^2 \|\tilde{\pi}_r - \pi_r\|_{S_2}^2 \leq 2\sigma^2 \lambda_1^2 \tilde{\Delta}_k =: \sigma_k^2,$$

where the last equality follows from the inclusion $\mathcal{C}_{C,i}(\tilde{\Delta}_k) \subset \mathcal{S}_{M,r}(2\tilde{\Delta}_k)$. By Lemma 8.2 and S_2 -chaining over $\mathcal{S}_{M,r}(2\tilde{\Delta}_k)$,

$$\begin{aligned} \mathbf{E} (0 \vee W_{k,i}^0) &\lesssim \int_0^1 \sigma \lambda_1 \tilde{\Delta}_k^{1/2} \left(2 \log \left(N(\mathcal{S}_{M,r}(\tilde{\Delta}_k), d_{S_2}, 2\tilde{\Delta}_k^{1/2} \delta) \right) \right)^{1/2} d\delta \\ &\lesssim \sigma \lambda_1 \tilde{\Delta}_k^{1/2} \sqrt{r(M-r)}. \end{aligned}$$

I.e., there exists some constant $c > 0$, independent of M, r, σ^2 and C , such that

$$(8.21) \quad \mathbf{E} (0 \vee W_{k,i}^0) \leq c \sigma \lambda_1 \tilde{\Delta}_k^{1/2} \sqrt{r(M-r)},$$

and by the Borell (1975) - Sudakov and Cirel'son (1974) inequality,

$$(8.22) \quad \mathbf{P} \left(W_{k,i}^0 \geq c \sigma \lambda_1 \tilde{\Delta}_k^{1/2} \sqrt{r(M-r)} + \sqrt{2} \sigma \lambda_1 \tilde{\Delta}_k^{1/2} \sqrt{2\eta} \right) \leq \exp(-\eta) \text{ for any } \eta > 0.$$

(THE CASE $i = 1$) With the help of (8.22) we evaluate the first term

$$\sum_{k < k_{0,1}} \mathbf{E} \left(0 \vee (W_{k,1}^0 - \Omega_{k,1}) \right) \leq \sum_{k < k_{0,1}} \mathbf{E} \left(0 \vee (W_{k,1}^0 - \frac{1}{2} \tilde{\Delta}_k \lambda_r^2) \right)$$

in expansion (8.20). For the ease of notation, define $\Delta_k := c \sigma \lambda_1 \tilde{\Delta}_k^{1/2} \sqrt{r(M-r)}$. Since Δ_k depends only linearly on $\tilde{\Delta}_k^{1/2}$, while

$$\|\pi_r C\|_{S_2}^2 - \|\tilde{\pi}_r C\|_{S_2}^2 \geq \frac{1}{2} \tilde{\Delta}_k \lambda_r^2 \text{ for all } \tilde{\pi}_r \in \mathcal{C}_C(\tilde{\Delta}_k) \text{ and } k \leq k_{0,1},$$

define the additional auxiliary integer

$$k_1^* := \arg \max_{k \leq k_{0,1}} \left\{ \frac{1}{2} \tilde{\Delta}_k \lambda_r^2 - \Delta_k \geq \frac{1}{4} \tilde{\Delta}_k \lambda_r^2 \right\},$$

and set $k_1^* = 0$ if the relation never holds. If $k_1^* = 0$, then

$$\sqrt{r} \lambda_r^2 \leq 4c \sigma \lambda_1 \sqrt{r(M-r)},$$

and the bound

$$\mathbf{E} \sup_{\tilde{\pi}_r \in \mathcal{S}_{M,r}} \left(\|\tilde{\pi}_r(C + \mathbf{E})\|_{S_2}^2 - \|\pi_r(C + \mathbf{E})\|_{S_2}^2 \right) \lesssim \sigma^2 r M \left(1 + \frac{\lambda_1}{\sigma \sqrt{M}} \right)$$

follows immediately by Corollary 4.2 (i). Thus, we assume $k_1^* \geq 1$. We first treat the case $k \leq k_1^*$. By the representation formula for the expectation of non-negative random variables and the definition of k_1^* ,

$$\begin{aligned} \mathbf{E}\left(0 \vee \left(W_{k,1}^0 - \frac{1}{2}\tilde{\Delta}_k\lambda_r^2\right)\right) &= \int_0^\infty \mathbf{P}\left(W_{k,1}^0 - \frac{1}{2}\tilde{\Delta}_k\lambda_r^2 > u\right)du \\ &\leq \int_0^\infty \mathbf{P}\left(Z_k^0 - \Delta_k > u + \frac{1}{4}\tilde{\Delta}_k\lambda_r^2\right)du \end{aligned}$$

for all $k \leq k_1^*$. Next, by (8.22) and with $A_k = (2\sigma_k)^{-1}\frac{1}{4}\tilde{\Delta}_k\lambda_r^2$,

$$\begin{aligned} \int_0^\infty \mathbf{P}\left(W_{k,1}^0 > u + \frac{1}{4}\tilde{\Delta}_k\lambda_r^2\right)du &\leq \int_0^\infty \exp\left(-\frac{\left(u + \frac{1}{4}\tilde{\Delta}_k\lambda_r^2\right)^2}{4\sigma_k^2}\right)du \\ (8.23) \quad &\lesssim \frac{\sigma_k}{2 + A_k} \exp(-A_k^2/2) \end{aligned}$$

$$(8.24) \quad \leq \frac{\sigma_k}{(2 + A_k)(1 + A_k^2/2)},$$

where we used $\mathbf{P}(N \geq x) \leq (2+x)^{-1}\exp(-x^2/2)$ for $N \sim \mathcal{N}(0,1)$ in (8.23) and the inequality $\exp(-x) \leq 1/(1+x) \forall x > 0$ in (8.24). Thus

$$(8.25) \quad \sum_{k \leq k_1^*} \mathbf{E}\left(0 \vee \left(W_{k,1}^0 - \Delta_k - \frac{1}{2}\tilde{\Delta}_k\lambda_r^2\right)\right) \lesssim \sum_{k \leq k_1^*} \left(\frac{\sigma_k}{1 + A_k^2/2}\right).$$

We evaluate (8.25). Recall $\sigma_k = \sqrt{2}\sigma\lambda_1\tilde{\Delta}_k^{1/2}$. Then, bounding the sum by an integral and by change of variables,

$$\begin{aligned} \sum_{k \leq k_1^*} \frac{\sigma_k}{1 + A_k^2/2} &\leq \sum_{k \in \mathbb{N}} \sqrt{2}\sigma\lambda_1\tilde{\Delta}_k^{1/2} \left(1 + \frac{\tilde{\Delta}_k^2\lambda_r^4}{128\sigma^2\lambda_1^2\tilde{\Delta}_k}\right)^{-1} \\ &\lesssim \sigma\lambda_1 \int_0^1 \left(1 + x^2 \frac{\lambda_r^4}{\sigma^2\lambda_1^2}\right)^{-1} dx \\ &\lesssim \sigma^2 \frac{\lambda_1^2}{\lambda_r^2} \int_0^\infty (1 + x^2)^{-1} dx. \end{aligned}$$

In order to estimate the expression $\sum_{k \geq k_1^*+1}^{k_{0,1}-1} \mathbf{E}(0 \vee (W_{k,1}^0 - (1/2)\tilde{\Delta}_k\lambda_r^2))$ we need to determine a lower bound on k_1^* in dependence of λ_1, λ_r . By its definition, $k > k_1^*$ implies

$$(8.26) \quad c\sigma\lambda_1\tilde{\Delta}_k^{1/2}\sqrt{r(M-r)} > \tilde{\Delta}_k\lambda_r^2/4$$

as long as $k_1^* < k_{0,1}$. Recalling by (8.21) that $\mathbf{E}(0 \vee W_{k,1}^0) \leq \Delta_k$, and using (8.26) and the representation for the tail of the geometric series,

$$\sum_{k > k_1^*}^{k_{0,1}-1} \mathbf{E}(0 \vee W_{k,1}^0) \lesssim \sum_{k > k_1^*} \tilde{\Delta}_k^{1/2}\sigma\lambda_1\sqrt{r(M-r)} \lesssim \sigma^2 r M \frac{\lambda_1^2}{\lambda_r^2}$$

As concerns the second term $\mathbf{E}\widetilde{W}_{k_{0,1}}$ in expansion (8.20), we obtain by definition (8.19) of $k_{0,1}$, if $k_{0,1} < \infty$,

$$\mathbf{E}\widetilde{W}_{k_{0,1}} \lesssim \widetilde{\Delta}_{k_{0,1}} \sigma \lambda_1 \sqrt{r(M-r)} \lesssim \left(\frac{\frac{1}{r} \sum_{i=r+1}^{2r} \lambda_i^2}{\lambda_r^2} \right)^{1/2} \cdot \sigma \lambda_1 r \sqrt{(M-r)}.$$

If $k_{0,1} = \infty$, then $\text{rank}(C) = r$ and the rank- r -projection π_r is unique, i.e. $\widetilde{W}_{k_{0,1}} = 0$. Collecting things together, this proves the bound

$$(8.27) \quad \delta_{C,M,\sigma^2,r} \lesssim \sigma^2 r M (\text{I} + \text{II}).$$

(THE CASE $i = 2$) We assume subsequently that $\lambda_r > \lambda_{r+1}$, because otherwise $\text{III} = \infty$ and the result follows with (8.27). We proceed similar to the case $i = 1$ above, but with $k_{0,2} := \infty$ and the auxiliary integer

$$k_2^* := \arg \max_{k \in \mathbb{N}} \left\{ (\lambda_r^2 - \lambda_{r+1}^2) \widetilde{\Delta}_k - \Delta_k \geq \frac{1}{2} (\lambda_r^2 - \lambda_{r+1}^2) \widetilde{\Delta}_k \right\},$$

where $k_2^* := 0$ if this relation never holds. The sum

$$\sum_{k < k_{0,2}} \mathbf{E} \left(0 \vee (W_{k,2}^0 - \Omega_{k,2}) \right) \leq \sum_{k \in \mathbb{N}} \mathbf{E} \left(0 \vee (W_{k,2}^0 - (\lambda_r^2 - \lambda_{r+1}^2) \widetilde{\Delta}_k) \right)$$

can be treated analogously to the case $i = 1$, with $\lambda_r^2 - \lambda_{r+1}^2$ in place of λ_r^2 . Similarly to (8.26), $k > k_2^*$ implies

$$c \sigma \lambda_1 \widetilde{\Delta}_k^{1/2} \sqrt{r(M-r)} > \widetilde{\Delta}_k (\lambda_r^2 - \lambda_{r+1}^2) / 2.$$

Since $\mathbf{E}(0 \vee W_{k,2}^0) \leq \Delta_k$, the representation for the tail of the geometric series yields, as above,

$$\sum_{k > k_2^*} \mathbf{E}(0 \vee W_{k,2}^0) \lesssim \sum_{k > k_2^*} \widetilde{\Delta}_k^{1/2} \sigma \lambda_1 \sqrt{r(M-r)} \lesssim \sigma^2 r M \frac{\lambda_1^2}{\lambda_r^2 - \lambda_{r+1}^2}.$$

Combining case (i) and (ii) yields the proof of the Theorem. \square

9. Proof of Theorem 6.2. Recall the definition $\widetilde{\Delta}_k := 2^{-k+2}r$, $k \in \mathbb{N}$. Since $\|\pi_s(\text{Id} - \widetilde{\pi}_s)\|_{S_2}^2 = \|(\text{Id} - \pi_s)\widetilde{\pi}_s\|_{S_2}^2$ and $\pi_s - \widetilde{\pi}_s = (\text{Id} - \pi_s)\widetilde{\pi}_s - \pi_s(\text{Id} - \widetilde{\pi}_s)$ is an orthogonal decomposition, observe first that

$$\begin{aligned} \|\pi_s C_{\alpha,s}\|_{S_2}^2 - \|\widetilde{\pi}_s C_{\alpha,s}\|_{S_2}^2 &= \alpha^2 \left(\|\pi_s\|_{S_2}^2 - \|\widetilde{\pi}_s \pi_s\|_{S_2}^2 \right) \\ &= \alpha^2 \|(\text{Id} - \widetilde{\pi}_s) \pi_s\|_{S_2}^2 \\ &= \alpha^2 \|\widetilde{\pi}_s(\text{Id} - \pi_s)\|_{S_2}^2 = \frac{\alpha^2}{2} \|\widetilde{\pi}_s - \pi_s\|_{S_2}^2, \end{aligned}$$

that is,

$$\mathbf{E}\|\pi_s(C_{\alpha,s} + \mathbf{E})\|_{S_2}^2 - \mathbf{E}\|\tilde{\pi}_s(C_{\alpha,s} + \mathbf{E})\|_{S_2}^2 \in (\alpha^2 \tilde{\Delta}_{k+1}, \alpha^2 \tilde{\Delta}_k] \Leftrightarrow \|\tilde{\pi}_s - \pi_s\|_{S_2}^2 \in (2\tilde{\Delta}_{k+1}, 2\tilde{\Delta}_k].$$

Recall the definition

$$\mathcal{G}_{M,s}(\delta, C) = \left\{ \tilde{\pi}_s \in \mathcal{S}_{M,s} : \|\pi_s C\|_{S_2}^2 - \|\tilde{\pi}_s C\|_{S_2}^2 \leq \delta \right\}.$$

Note at this point that with $\bar{C}_{\alpha, M-s} := \alpha U(\text{Id} - \text{Id}_s)V'$, we have $\tilde{\pi}_r \in \mathcal{G}_{M,r}(\delta, C_{\alpha,r}) \Leftrightarrow (\text{Id} - \tilde{\pi}_r) \in \mathcal{G}_{M, M-r}(\delta, \bar{C}_{\alpha, M-r})$. Define

$$k^{**} := \arg \max_{k \in \mathbb{N}} \left\{ \alpha^2 \tilde{\Delta}_k \geq d\sigma^2 s(M-s) \right\}$$

for some $d > 0$ to be specified later, and let

$$\mathcal{D}_s(\alpha) := \mathcal{G}_{M,s}(\alpha^2 \tilde{\Delta}_{k^{**}+1}, C_{\alpha,s}) \quad \text{as well as} \quad \bar{\mathcal{D}}_s(\alpha) := \mathcal{G}_{M,s}(\alpha^2 \tilde{\Delta}_{k^{**}+1}, \bar{C}_{\alpha,s}).$$

Note that $k^{**} \rightarrow \infty$ as $\alpha \rightarrow \infty$. It holds that

$$\begin{aligned} & \mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{S}_{M,s}} \left\| \tilde{\pi}_s(C_{\alpha,s} + \mathbf{E}) \right\|_{S_2}^2 - \left\| \pi_s(C_{\alpha,s} + \mathbf{E}) \right\|_{S_2}^2 \right) \\ & \geq \mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \left\| \tilde{\pi}_s \mathbf{E} \right\|_{S_2}^2 - \left\| \pi_s \mathbf{E} \right\|_{S_2}^2 + 2\alpha \text{tr} \left(\mathbf{E}'(\tilde{\pi}_s - \pi_s)\pi_s \right) - ds(M-s)\sigma^2 \right) \\ & \geq \mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} 2\alpha \text{tr} \left(\mathbf{E}'(\tilde{\pi}_s - \pi_s)\pi_s \right) - ds(M-s)\sigma^2 \right) \\ & \quad - \mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \left\| \pi_s \mathbf{E} \right\|_{S_2}^2 - \left\| \tilde{\pi}_s \mathbf{E} \right\|_{S_2}^2 \right). \end{aligned}$$

Because of $\limsup_{\alpha \rightarrow \infty} \mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \left\| \pi_s \mathbf{E} \right\|_{S_2}^2 - \left\| \tilde{\pi}_s \mathbf{E} \right\|_{S_2}^2 \right) = 0$, it remains to prove that

$$\begin{aligned} & \liminf_{\alpha \rightarrow \infty} \max_{s \in \{r, M-r\}} \mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} 2\alpha \text{tr} \left(\mathbf{E}'(\tilde{\pi}_s - \pi_s)\pi_s \right) - ds(M-s)\sigma^2 \right) \\ (9.1) \quad & = \sigma^2 \liminf_{\alpha \rightarrow \infty} \max_{s \in \{r, M-r\}} \mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} 2(\alpha/\sigma) \text{tr} \left((\mathbf{E}/\sigma)' \tilde{\pi}_s \pi_s \right) - ds(M-s) \right) \\ & \gtrsim \sigma^2 r(M-r). \end{aligned}$$

First, we have

$$\begin{aligned} & \mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \text{tr} \left((\mathbf{E}/\sigma)' \tilde{\pi}_s \pi_s \right) \\ & = \mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \text{tr} \left((\mathbf{E}/\sigma)' (\text{Id} - \tilde{\pi}_s) \pi_s \right) \\ & = \mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \left(\text{tr} \left((\mathbf{E}/\sigma)' \tilde{\pi}_s \right) - \text{tr} \left((\mathbf{E}/\sigma)' \tilde{\pi}_s (\text{Id} - \pi_s) \right) \right), \end{aligned}$$

which implies

$$(9.2) \quad \mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'(\operatorname{Id} - \tilde{\pi}_s)\pi_s) + \mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_s(\operatorname{Id} - \pi_s)) \\ \geq \mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_s).$$

In case $M \in 2\mathbb{N}$ and $s = M/2$, since $\|\pi_s - \tilde{\pi}_s\|_{S_2}^2 = \|(\operatorname{Id} - \pi_s) - (\operatorname{Id} - \tilde{\pi}_s)\|_{S_2}^2$, both expectations on the LHS in the inequality (9.2) are identical for reasons of symmetry, which leads to

$$\mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_s\pi_s) \geq \frac{1}{2} \mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_s) \quad \text{in case } s = M/2.$$

Although the polar decomposition of $(\operatorname{Id} - \tilde{\pi}_s)\pi_s$ and $\tilde{\pi}_s(\operatorname{Id} - \pi_s)$, respectively, suggests a similar symmetry argument, we do not have a rigorous treatment of an argument of this type yet, and remain therefore with the inequality

$$\max_{s \in \{r, M-r\}} \mathbf{E} \sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'(\operatorname{Id} - \tilde{\pi}_s)\pi_s) \geq \frac{1}{2} \mathbf{E} \sup_{\tilde{\pi}_r \in \mathcal{D}_r(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_r)$$

only. Note at this point that

$$\begin{aligned} \mathbf{E} \sup_{\tilde{\pi}_r \in \mathcal{D}_r(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_r) &= \mathbf{E} \sup_{\tilde{\pi}_{M-r} \in \tilde{\mathcal{D}}_{M-r}(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_{M-r}) \\ &= \mathbf{E} \sup_{\tilde{\pi}_{M-r} \in \mathcal{D}_{M-r}(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_{M-r}), \end{aligned}$$

where the second equality follows from invariance of the above expression under orthogonal transformation. By Sudakov's minoration and the bound (8.13) of Lemma 8.2,

$$\mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_s) \right) \gtrsim \delta \sqrt{\log N(\mathcal{D}_r(\alpha), d_{S_2}, \delta)} \geq \delta \sqrt{r(M-r)} \sqrt{\log \left(\frac{c\tilde{\Delta}_{k^{**}}^{1/2}}{\sqrt{2}\delta} \right)}$$

for any arbitrary $0 < \delta < c\tilde{\Delta}_{k^{**}}^{1/2}/\sqrt{2}$, $s \in \{r, M-r\}$, where we used that

$$\mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{D}_r(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'(\operatorname{Id} - \tilde{\pi}_r)) \right) = \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{D}_r(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_r) \right)$$

and

$$N(\mathcal{D}_r(\alpha), d_{S_2}, \delta) \geq \left(\frac{c\tilde{\Delta}_{k^{**}}^{1/2}}{\sqrt{2}\delta} \right)^{r(M-r)}$$

with the constant c of Lemma 8.2. The choice $\delta = c\tilde{\Delta}_{k^{**}}^{1/2}/8$ yields finally

$$(9.3) \quad \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{D}_r(\alpha)} \operatorname{tr}((\mathbf{E}/\sigma)'\tilde{\pi}_r) \right) \geq K\tilde{\Delta}_{k^{**}}^{1/2} \sqrt{r(M-r)}$$

for some constant $K > 0$ which does not depend on σ , M and α . Thus,

$$(9.1) = \sigma^2 \liminf_{\alpha \rightarrow \infty} \max_{s \in \{r, M-r\}} \mathbf{E} \left(\sup_{\tilde{\pi}_s \in \mathcal{D}_s(\alpha)} 2(\alpha/\sigma) \operatorname{tr}((\mathbf{E}/\sigma)' \tilde{\pi}_s \pi_s) - dr(M-r) \right)$$

$$(9.4) \geq \sigma^2 \liminf_{\alpha \rightarrow \infty} \mathbf{E} \left(\sup_{\tilde{\pi}_r \in \mathcal{D}_r(\alpha)} (\alpha/\sigma) \operatorname{tr}((\mathbf{E}/\sigma)' \tilde{\pi}_r) - dr(M-r) \right).$$

Choosing now d in the definition of k^{**} largest possible such that $K\sqrt{d}-d \geq K\sqrt{d}/2$ and plugging the lower bound (9.3) into (9.4) proves the Theorem. \square

References.

- Artiles, L., Gill, R., and Guta, M. (2004). An invitation to quantum tomography. *J. Royal Statistical Society, Series B*(67):109–134.
- Bai, Z. and Yin, Y. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.*, 21:1275–1294.
- Bechar, I. (2009). A Bernstein type inequality for stochastic processes of quadratic forms of Gaussian variables. <http://arxiv.org/pdf/0909.3595>.
- Borell, C. (1975). The Brunn-Minkowski inequality in Gauß space. *Inv. Math.*, 30:207–216.
- Bunea, F., She, Y., and Wegkamp, M. (2011). Optimal selection of reduced-rank estimators of high-dimensional matrices. *Ann. Statist.*, 39(2):1282–1309.
- Capitaine, M., Donati-Martin, C., and Féral, D. (2009). The largest eigenvalue of finite rank deformations of large Wigner matrices: convergence and non universality of the fluctuations. *Ann. Prob.*, 37(1):1–47.
- Capitaine, M., Donati-Martin, C., and Féral, D. (2012). Central limit theorems for eigenvalues of deformations of Wigner matrices. *Ann. Inst. H. P.*, to appear.
- Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge Studies in advanced mathematics. Cambridge University Press, Cambridge, UK.
- Johnson, W. and Lindenstrauss, J. (2001). *Handbook of the Geometry of Banach Spaces*. Elsevier Science B.V., Volume 1.
- Koltchinskii, V. (2011). Von Neumann entropy penalization and low rank matrix estimation. *Ann. Statist.*, 39:2936–2973.
- Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28:1302–1338.
- Ledoux, M. (1996). *Isoperimetry and Gaussian Analysis*. Ecole d’été de Probabilités St.-Flour 1994. Lecture notes in Mathematics 1684. Springer.
- Negahban, S. and Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Statist.*, 39:1069–1097.
- Pajor, A. (1998). Metric entropy of the Grassmann manifold. *Convex Geometric Analysis*, 34:181–188.
- Pizzo, A., Renfrew, D., and Soshnikov, A. (2012). On finite rank deformations of Wigner matrices. *Ann. Inst. H. P.*, to appear.
- Rohde, A. and Tsybakov, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Ann. Statist.*, 39:887–930.
- Sudakov, V. and Cirel’son, C. (1974). Extremal properties of half-spaces for spherically invariant measures. (Russian). *Zap. Nauch. Sem. Leningrad. Otdel. Mat. Inst. Steklov.*, 41:14–24.

- Szarek, S. (1982). Nets of Grassmann manifold and orthogonal groups, University of Iowa Press. *Proceedings of Banach Spaces Workshop*, pages 169–185.
- Talagrand, M. (1992). Sudakov-type minoration for Gaussian chaos processes. *Israel Journal of Mathematics*, 79:207–224.
- Tao, T. (2012). Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory and Relat. Fields*, to appear.
- van der Vaart, A. W. (1996). New Donsker classes. *Ann. Probab.*, 24(4):2128–2140.

UNIVERSITÄT HAMBURG, FACHBEREICH MATHEMATIK, BUNDESSTRASSE 55, 20146 HAMBURG, GERMANY
E-MAIL: angelika.rohde@math.uni-hamburg.de